

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection School Of Economics

School of Economics

10-2011

Non-parametric Regression under Location Shifts

Peter C. B. PHILLIPS

Singapore Management University, peterphillips@smu.edu.sg

Liangjun SU

Singapore Management University, ljsu@smu.edu.sg

DOI: <https://doi.org/10.1111/j.1368-423X.2011.00344.x>

Follow this and additional works at: https://ink.library.smu.edu.sg/soe_research



Part of the [Econometrics Commons](#)

Citation

PHILLIPS, Peter C. B. and SU, Liangjun. Non-parametric Regression under Location Shifts. (2011). *Econometrics Journal*. 14, (3), 457-486. Research Collection School Of Economics.

Available at: https://ink.library.smu.edu.sg/soe_research/1364

This Journal Article is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Nonparametric Regression under Location Shifts*

Peter C. B. Phillips

Yale University, University of Auckland

University of Southampton & Singapore Management University

Liangjun Su

School of Economics

Singapore Management University

May 11, 2010

Abstract

Recent work by Wang and Phillips (2009b, 2010) has shown that ill posed inverse problems do not arise in nonstationary nonparametric regression and there is no need for nonparametric instrumental variable estimation. Instead, simple Nadaraya Watson nonparametric estimation of a (possibly nonlinear) cointegrating regression equation is consistent with a limiting (mixed) normal distribution irrespective of the endogeneity in the regressor, near integration as well as integration in the regressor, and serial dependence in the regression equation. The present paper shows that some closely related results apply in the case of structural nonparametric regression with independent data when there are continuous location shifts in the regressor. In such cases, location shifts serve as an instrumental variable in tracing out the regression line similar to the random wandering nature of the regressor in a cointegrating regression. Asymptotic theory is given here for local level and local linear nonparametric estimators and links with nonstationary cointegrating regression theory and nonparametric IV regression are explored.

Some interesting cases are discovered where nonparametric regression is consistent whereas parametric regression is inconsistent even when the true regression functional form is known and used in regression. This appears to be a paradox, as knowing the true functional form should not in general be detrimental in regression. In the present case, local regression methods turn out to

*Phillips acknowledges partial support from NSF Grant #SES 06-47086. Address correspondence to Peter C. B. Phillips, Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, Connecticut USA 06520-8281; e-mail: peter.phillips@yale.edu. Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; e-mail: ljsu@smu.edu.sg.

have a distinct advantage because of endogeneity in the regressor. The paradox arises because additional correct information is not necessarily advantageous when information is incomplete. In the present case, endogeneity in the regressor introduces bias when the true functional form is known, but interestingly does not do so in local nonparametric regression. We propose two new consistent estimators for the parametric regression, which address the endogeneity in the regressor by means of spatial bounding and bias correction using nonparametric estimation.

JEL classification: C13, C14.

Keywords: Bias-correction, Endogeneity, Kernel regression, Location shift, Nonparametric IV, Nonstationarity, Paradox, Spatial \mathcal{L}_2 regression, Structural estimation.

1 Introduction and Motivation

Much recent interest in nonparametric regression has focused on the effects of endogeneity in the regressor. Attention has primarily been microeconomic and so has naturally concentrated on a framework in which the observed data are independently distributed. Allowing for endogeneity is particularly important in practical applications where there are unobserved characteristics (such as inherent ability) that influence both the regressors and the equation errors. In such contexts, the unknown regression function is not recoverable as a conditional expectation but is submerged in a functional integral equation whose solution raises difficulties that fall under the category of ill-posed inversion problems. Research on this type of problem in econometrics has been underway for about a decade, following on from a much longer literature in mathematics, numerical analysis, image processing, and statistics. Methods of “regularizing” the inversion problem have become popular in the theoretical development of the subject and both kernel and series-based approaches have been considered. There is now a wide literature on the subject - see Florens (2003) and Carrasco, Florens, and Renault (2007) for recent overviews. In addition to microeconomic contexts, the field of potential applications has recently widened to include asset pricing in financial econometrics (Simoni, 2009).

To facilitate the use of standard functional analysis methods, it has become conventional in econometric treatments to restrict the system variables to have bounded support and bounded densities. These restrictions appear innocent given that (distribution function) transformations can mechanically transform the support of all variables to the $[0, 1]$ interval. However, these transformations can induce subtle changes in the system (such as unbounded densities) that are not innocuous and they can omit factors that are relevant in microeconomic modeling. For example, in applied microeconomics, adding more data generally means adding more parameters to estimate and introducing more variation to be explained. Indeed, those very characteristics have driven much of the research on robust estimation and the treatment of individual fixed effects.

Some of these characteristics manifest in an important way in the setting explored in the present paper. In particular, we show that the variation induced by locational shifts in the regressor can have an enormous impact on the potential capabilities of nonparametric regression, completely removing the ill posed inverse problem and facilitating consistent estimation by conventional nonparametric regression techniques.

Fixed location effects may arise in various economic contexts. For example, markets for a product may be differentiated by location with different supply functions where the supply curve is influenced by factors that shift the supply curve around according to location. Such shifts in a supply curve are precisely the ones that are described in textbook treatments of identification in simultaneous equations models of supply and demand. As the supply curve moves around it traces out a sequence of equilibria associated with the Marshallian cross at each location.

The key idea is illustrated in Fig. 1, which provides a sample plot of data generated from the following location specific linear Marshallian stochastic demand/supply system

$$\text{Demand:} \quad q_i = a + bp_i + u_i, \quad (1.1)$$

$$\text{Supply:} \quad p_i = c + \sum_{\alpha=1}^K \mu_{\alpha} 1\{i \in A_{\alpha}\} + u_{pi}, \quad (1.2)$$

$$u_i = \{\theta u_{pi} + \epsilon_{qi}\} / (1 + \theta^2)^{1/2},$$

where $1\{\cdot\}$ is the usual indicator function, A_{α} 's are disjoint clusters of individuals associated with a mean price $c + \mu_{\alpha}$, the errors $(u_{pi}, \epsilon_{qi}) \equiv iidN(0, I_2)$, and α signifies a particular market location in which, if there were no disturbances, the product price would be $p = c + \mu_{\alpha}$ and demand would be $q = a + b(c + \mu_{\alpha})$. In general, the locational equilibrium at location α is disturbed by errors and the data tend to cluster around each locational equilibrium point. The nature and orientation of the clusters depends on the joint distribution of the equation errors (u_i, u_{pi}) , which in turn depends on the endogeneity parameter θ , corresponding to which we have the correlation $\rho = \text{corr}(u_i, u_{pi}) = \theta / (1 + \theta^2)^{1/2}$. When $\theta = 0$, there is no endogeneity ($\rho = 0$), and when $\theta = 2$, there is substantial endogeneity ($\rho = 0.89$). Fig. 1 illustrates this locational clustering phenomena with a typical data set for $K = 5$ locations corresponding to $\mu_{\alpha} \in \{-4, -2, 0, 2, 4\}$ with $\theta = 2$, $c = 5$, $a = 10$, $b = -1$, and $M = 100$ observations for each α . Along the demand curve we observe clusters of points around price levels $\{1, 3, 5, 7, 9\}$ corresponding to each of the market locations. As the location α shifts, the data display a tendency to trace out the demand curve, just as in textbook discussions.

Fig. 2 displays the Nadaraya Watson local level estimate of the (demand) curve using all $n = M \cdot K = 500$ observations, a Gaussian kernel and a mechanical $n^{-1/5}$ bandwidth rule. As is apparent, within the support of the sample data for (q_i, p_i) the regression line is very well fitted by the nonparametric curve irrespective of the endogeneity. In this region, the data trace out the demand curve and it is this line that is well approximated by the regression. But

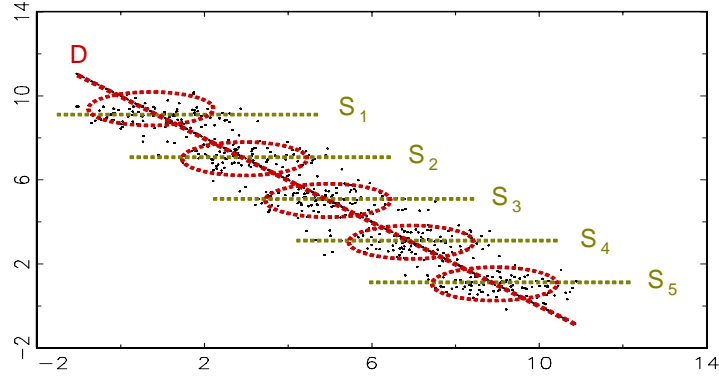


Figure 1: Demand and supply curves (for 5 locations) with sample data generated by (1.1) and (1.2).

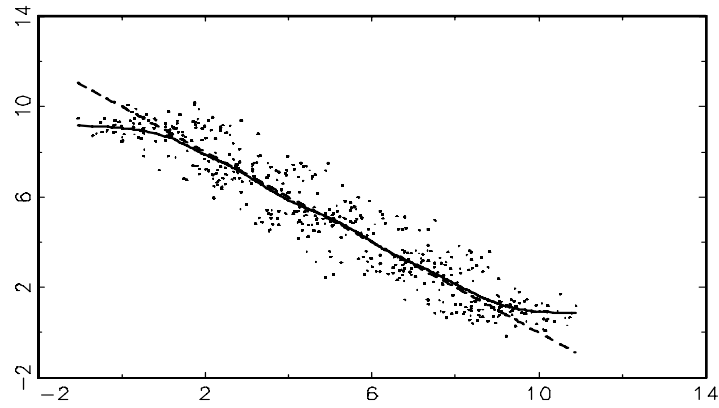


Figure 2: Nadaraya Watson nonparametric regression estimate of the demand curve (1.1) using sample data taken over all locations.

at the extremes of the support the endogeneity in the data is more clearly manifest and the regression line tracks out the data, following the supply correspondence in those two regions. Note that in this example only five discrete location shifts are needed in order to achieve this high level of accuracy in nonparametric regression.

This simple example illustrates the first central finding of the present paper. In spite of endogeneity in a regressor and provided the regressor has enough variation, simple nonparametric regression may be consistent even with independent data, at least within the interior of the support. The simplest manner in which sufficient variation may be attained is for the regressors to sustain location shifts by way of external fixed effects, as in the illustration. These location shifts serve the role of a form of cross section nonstationarity and in this sense the resulting consistency of nonparametric regression is analogous to that achieved in the case of a (near) unit root regressor. In that case, the regressor is recurrent and visits every point in the space an infinite number of times, which is broadly speaking analogous to a regressor which undergoes continuous shifts in location as data accumulate, so that the variation of the regressor continues to increase with the sample size. Wang and Phillips (2009b, 2010) showed that nonparametric regression with unit root or near integrated regressors is consistent, but with a reduced rate of convergence compared with the conventional \sqrt{nh} rate for sample size n and bandwidth h . Correspondingly, the present paper shows that within the support and with a reduced rate of convergence, nonparametric regression is consistent at all points where the location shifts are continuous in the limit as $n \rightarrow \infty$.

The case where the distribution of the regressor has compact support is also investigated. As mentioned above, this case has been extensively studied in the nonparametric IV literature. In the present context, we show that the nonparametric local level and local linear estimators can both achieve the conventional \sqrt{nh} convergence rate of nonparametric estimation in models with exogenous regressors. However, parametric estimation turns out to be inconsistent, even though the true form of the (parametric) regression function is known and estimated. This outcome is paradoxical, as local level regression is consistent whereas linear parametric regression that uses the true functional form and global information is inconsistent. Clearly, using all information is costly here. The reason, of course, is the endogeneity in the regressor. Intriguingly, however, partial information usage of the type employed in kernel regression avoids most of the problems associated with parametric regression.

To illustrate this phenomenon, we generate a sample of $n = 500$ observations from the following system

$$\begin{aligned} y_i &= \beta_0 + \beta_1 X_i + u_i, \quad \beta_0 = 10, \quad \beta_1 = -1, \\ X_i &= \sum_{\alpha=1}^5 \mu_\alpha 1\{i \in A_\alpha\} + u_{xi}, \\ u_i &= 2(\epsilon_{yi} + \gamma u_{xi}) / (1 + \gamma^2)^{1/2}, \end{aligned}$$

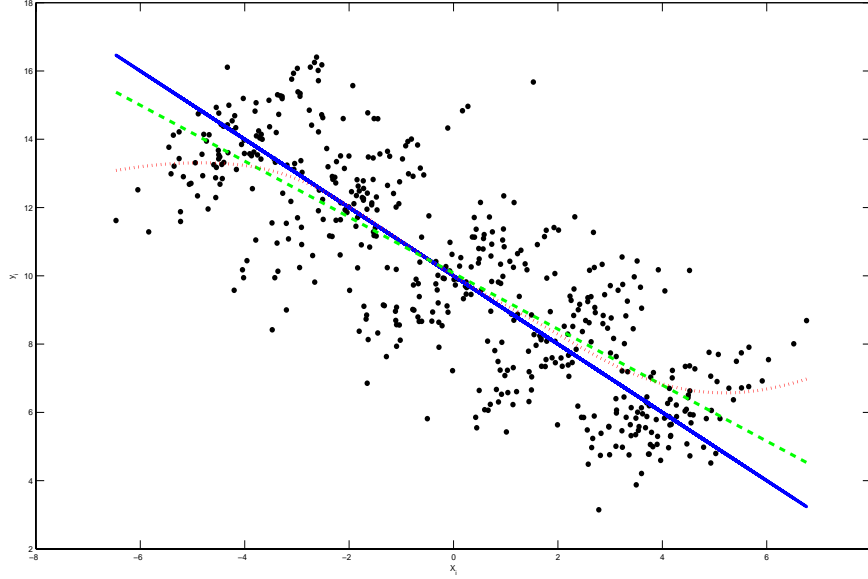


Figure 3: Local level nonparametric (dotted) and linear regression (dashed) estimates of the true regression line (solid) using the full sample (scatter plot) over all locations.

where $\gamma = 2.07$, (ϵ_{yi}, u_{xi}) are *iid* $N(0, I_2)$, and $\mu_\alpha \in \{-4, -2, 0, 2, 4\}$. When $\gamma = 2.07$, the error correlation coefficient is $\text{corr}(u_i, u_{xi}) = 0.9$, implying strong endogeneity in the structural equation. Fig. 3 provides a typical sample plot of data generated from this system. Also displayed are the true linear regression line, the local level nonparametric estimate (using a Gaussian kernel and Silverman's rule of thumb for the bandwidth choice) and the fitted OLS estimate obtained for these data. Clearly, the local level estimate considerably outperforms the parametric estimate of the regression line over a wide range of the support of the regressor. But at the tails of the support the endogeneity bias becomes manifest and the location shifts lose power in identifying the regression line. Since it uses local information and does so increasingly as the sample size rises, the local level estimator at interior points of the domain of the regressor very effectively attenuates distortion from tail observations. On the other hand, parametric linear regression, which treats all observations as equally important and applies global information in fitting the regression line, is inevitably subject to potential distortions from tail observations.

This example illustrates the second central finding of the present paper. It makes clear that local nonparametric regression has robustness advantages beyond robustness to specific functional form, for which it is most commonly celebrated. As shown here, nonparametric regression may also display a robustness to endogeneity in a regression by concentrating attention on local information and attenuating tail information that may be more heavily subjected to endogeneity effects.

Intuitively, any regression approach like conventional parametric regression that uses global information can be subject to distortionary effects from outlying observations. Such behavior is very well known in statistics. Anscombe (1960) coined the term ‘outlier’, provided a brief history of the subject, and suggested trimming techniques to attenuate the effects of outliers in regression based on the insurance analogy of ‘protection and premium’ to guard against unwanted effects. In the present context, the reason for the outlier effect is that at the limits of the domain of definition, the observations are more affected by the endogeneity of X_i , so bias arising from the ends of the domain can dominate a global regression and result in inconsistency. By contrast, in nonparametric regression mainly local information is used in estimation so the endogeneity effects of X_i in the tail can be well controlled and first order bias, at least, can be eliminated in local regression. In effect, by concentrating attention on the cluster of observations around each point, nonparametric regression localizes attention and removes outlier effects. This heuristic suggests that to recover the true regression line by a parametric method, a natural approach is to modify the regression by removing the effects of tail observations. The idea is comparable to that of trimming or Winsorizing the data, on which there is a large literature in statistics stemming largely from Anscombe’s (1960) study (e.g. Welsh, 1987; Chen, Welsh and Chan, 2001). In the present context, to use Anscombe’s analogy, the idea is to provide protection (against the possible effects of endogeneity) by paying a premium in terms of losing some observations. Kernel regression accomplishes this task by using data that is effectively in the locality of each individual regression point, thereby sacrificing an (asymptotically larger) infinity of observations to achieve a local regression fit and, incidentally in the process, protection from the effects of endogeneity.

Continuous location shifts are idealized in a cross section setting because practical empirical examples with cross section data where continuous location shifts may occur are likely to be uncommon in economics. However, as the above illustration shows even a small number of discrete location shifts is sufficient to give nonparametric regression a significant advantage in dealing with endogeneity over a subset of the support. In addition, panel models that involve individual effects covering a large and continuously distributed population present an interesting and potentially realistic applications of the results given here. In particular, Phillips and Su (2009) discover that a structural nonparametric panel data model with individual effects covering a large population provide a natural mechanism for identifying structural nonparametric elements. In this case, the dispersion of the individual effects in the population introduces a form of leverage similar to that of continuous location shifts in the cross section case, thereby helping to trace out functional form and enable consistent estimation even in the presence of endogenous regressors.

The present paper introduces and explores this topic. Section 2 lays out a formal model and presents the main finding on the consistency and limiting normal distribution of nonparametric regression in a structural model with continuous location shifts. A series of remarks on the main

theorem then discuss the significance, operational import, and various extensions of the result. The discussion makes clear that a central requirement in achieving consistency and avoiding the use of external instrumentation is the continuity in the limit of the location shifts, although as the above illustration shows even a small number of discrete shifts can achieve a great deal of bias reduction in practice. Particular consideration is given to whether the regressor has compact support, as is frequently assumed in the latter literature, or infinite support, as in the nonstationary time series literature. The case of compact support is especially interesting because of the differences between parametric and nonparametric regression.

Section 3 studies the asymptotic properties of the OLS estimator for the coefficients in a linear model with an endogenous regressor that is generated via location shifts. We demonstrate the inconsistency of the OLS estimator in case of compact support and derive its asymptotic distribution. Even when the true functional form is known, parametric regression is generally inconsistent because of endogeneity. In the case of infinite support, even if parametric regression is consistent, its convergence rate can be arbitrarily slow in relation to nonparametric estimation. For this reason, we propose two new consistent estimators. The first is a spatial \mathcal{L}_2 estimator and the second is a bias-corrected OLS estimator that is based on spatial \mathcal{L}_2 regression residuals. For both cases, we show that the resulting estimators are consistent and asymptotically normally distributed. Section 4 concludes. Proofs of the main results are given in Section 5.

2 Limit theory under location shifts

2.1 Models with infinite support

This section introduces a nonparametric regression model in which there are continuous location shifts in the regressor whose support is infinite. Observations $\{(y_i, X_i), 1 \leq i \leq n\}$ are generated as follows:

$$y_i = g(X_i) + u_i, \quad E(u_i|X_i) \neq 0 \quad (2.1)$$

$$X_i = \sum_{\alpha=-m}^m \mu_\alpha 1\{i \in A_\alpha\} + u_{xi}, \quad (2.2)$$

$$\mu_\alpha = \frac{\alpha L_n}{2m}, \quad \alpha = -m, -m+1, \dots, m, \quad (2.3)$$

where $\{A_\alpha\}$ are disjoint clusters of individuals associated with locations $\{\mu_\alpha\}$ for the endogenous regressor X , and both $m \equiv m_n$ and L_n depend on the sample size n . Let $M \equiv M_n = \sum_{i=1}^n 1\{i \in A_\alpha\}$. The mechanism allows for M observations in the vicinity of each locality and may be generalized to allow M to depend on α , although that extension involves more notational complexity so it will not be pursued here. The total observation count is then $n = (2m+1)M$. Throughout the paper we require that as $n \rightarrow \infty$, $m \rightarrow \infty$, but M can either

be finite or pass to ∞ . In the case where X has infinite support we require $L_n \rightarrow \infty$ as $n \rightarrow \infty$. Otherwise, L_n can either be finite or pass to ∞ .

We start by considering the local constant (Nadaraya-Watson) kernel estimator of $g(x)$

$$\hat{g}(x) = \frac{\frac{1}{n} \sum_{i=1}^n y_i K_h(X_i - x)}{\frac{1}{n} \sum_{i=1}^n K_h(X_i - x)}, \quad (2.4)$$

where $K(\cdot)$ is a kernel function, $K_h(\cdot) = h^{-1}K(\cdot/h)$ and $h \equiv h_n$ is a bandwidth parameter. We make the following assumptions.

Assumption A1. $(u_i, u_{xi}), i = 1, \dots, n$, are independent and identically distributed (iid).

Assumption A2. $E(u_i) = 0$, $E(u_i^2) = \sigma^2$, and $E|u_i|^{2+\delta} < \infty$ for some $\delta > 0$.

Assumption A3. (i) The probability density function (p.d.f.) $f(\cdot, \cdot)$ of (u_i, u_{xi}) exists. $f(\cdot, \cdot)$ has second order partial derivative $f_2''(u, u_x)$ with respect to u_x such that $f_2''(u, u_x)$ is continuous in u_x and $\int \int |u f_2''(u, u_x)| du du_x < \infty$. The marginal p.d.f. of u_{xi} , $f_{u_x}(\cdot)$, has second order continuous derivatives such that $\int_{-\infty}^{\infty} |f_{u_x}'(p)| dp < \infty$, and $\int_{-\infty}^{\infty} |f_{u_x}''(p)| dp < \infty$.
(ii) There exists a function $C_f(\cdot)$ such that for any sequence $p_n \rightarrow \infty$, we have $|\int_{-\infty}^{-p_n} f(u, u_x) du_x + \int_{p_n}^{\infty} f(u, u_x) du_x| \leq p_n^{-v} C_f(u)$ for some $v > 1$ and $\int_{-\infty}^{\infty} |u| C_f(u) du < \infty$.
(iii) $f_{u_x}(p_n) = O(|p_n|^{-v-1})$ as $|p_n| \rightarrow \infty$.

Assumption A4. For given x , $g(x)$ has a continuous, bounded second derivative in a small neighborhood of x .

Assumption A5. The kernel function $K(\cdot)$ is a uniformly bounded symmetric p.d.f. such that $\int x^4 K(x) dx < \infty$.

Assumption A6. As $n \rightarrow \infty$, $m \rightarrow \infty$, $L_n \rightarrow \infty$, $h \rightarrow 0$, $nh/L_n \rightarrow \infty$, $nh^5/L_n \rightarrow c \in [0, \infty)$, and $nh(L_n^3 m^{-4} + L_n^{-2v-1}) \rightarrow 0$.

The iid condition in Assumption A1 can be relaxed. For example, we can allow the process $\{(u_i, u_{xi}), i \geq 1\}$ to be strictly stationary and strong mixing with mixing coefficients that decay to zero at certain rates. Alternatively, it is possible to allow for unit root or near unit root behavior in u_{xi} , as in Wang and Phillips (2009b, c). Assumption A2 is weak. The zero-mean condition is needed for the identification of the nonparametric function $g(\cdot)$. But we do not impose the exogeneity condition $E(u_i|X_i) = 0$ a.s., nor do we assume conditional homoskedasticity. Assumption A3 imposes some smoothness and tail conditions on the joint and marginal p.d.f.'s. The tail condition on $f_{u_x}(\cdot)$ is equivalent to requiring that the first moment of u_{xi} exists. Assumption A4 imposes a standard smoothness condition on the regression function $g(\cdot)$. Assumption A5 is also standard, although the symmetry of the kernel function facilitates analysis and simplifies notation. Assumption A6 imposes conditions on the bandwidth (h) and its association with n , m , L_n , and v .

We now state the first main result.

Theorem 2.1 *Under Assumptions A1-A6, we have*

$$\sqrt{nh/L_n} \left\{ \hat{g}(x) - g(x) - \frac{1}{2} h^2 \mu_2(K) g''(x) \right\} \rightarrow_d N(0, \sigma^2 \nu_2(K)), \quad (2.5)$$

where $\mu_2(K) = \int x^2 K(x) dx$, and $\nu_2(K) = \int K(x)^2 dx$. If in addition $c = 0$ in Assumption A6, then the bias term in the braces vanishes.

The following remarks discuss this theorem in connection with limit theory in other cases, including cases where there are no location shifts, where there is exogeneity, and where the regressor is integrated.

Remark 1. (Comparison with the exogenous regressor case) When the regressor X_i is exogenous, stationary and has p.d.f. $f(x)$, the local constant estimator has the following limit behavior

$$\begin{aligned} \sqrt{nh} \left(\hat{g}(x) - g(x) - h^2 \mu_2(K) \left\{ g'(x) f'(x) / f(x) + \frac{1}{2} g''(x) \right\} \right) \\ \rightarrow_d N(0, \sigma^2(x) \nu_2(K) / f(x)), \end{aligned} \quad (2.6)$$

where $\sigma^2(x) = E(u_i^2 | X_i = x)$. While this result is obviously similar to the limit theory in (2.5) there are some important distinctions. The most important difference is that the expression for the bias in (2.5) involves only a single term, which corresponds to the second bias term in (2.6). The analog of the first bias term in (2.6) in the present setting is

$$h^2 \mu_2(K) g'(x) \int_{-\infty}^{\infty} f'_{u_x}(p) dp,$$

which is 0 as the marginal p.d.f. $f_{u_x}(\cdot)$ vanishes at infinity. This explains why there is no linear term in the bias function appearing in (2.5), in contrast to the limit theory for local level estimation with stationary regressors. In the present context, endogeneity does not prevent consistency and makes a smaller $o(h^2)$ contribution to the bias so it does not figure in the limit theory (2.5). A second important distinction is that the marginal density of X_i appears in (2.6) but not in (2.5). The reason is that the denominator in the definition of $\hat{g}(x)$

$$\frac{1}{n} \sum_{i=1}^n K_h(X_i - x)$$

does not converge in probability to a density. Instead, as shown in the appendix,

$$\frac{L_n}{n} \sum_{i=1}^n K_h(X_i - x) \rightarrow_p 1,$$

and this result arises because the location shifts in X_i imply that the averaging operator $(L_n/n) \sum_{i=1}^n$ ensures that the density of u_{xi} is averaged over the whole domain leading to

$\int f_{u_x}(x) dx = 1$. A third important distinction is that the (unconditional) variance of u_i appears in (2.5) whereas the conditional variance $\sigma^2(x)$ of u_i appears in (2.6). The reason is that in the bias-variance decomposition of $\hat{g}(x) - g(x)$, the variance term

$$\frac{\sqrt{h}}{\sqrt{n}} \sum_{i=1}^n u_i K_h(X_i - x)$$

does not converge weakly to a normal distribution with mean zero and variance $\sigma^2(x) \nu_2(K) / f(x)$. Instead, the presence of location shifts in X_i leads to further averaging in the central limit theory and we have

$$\frac{\sqrt{hL_n}}{\sqrt{n}} \sum_{i=1}^n \{u_i K_h(X_i - x) - E[u_i K_h(X_i - x)]\} \rightarrow_d N(0, \sigma^2 \nu_2(K)).$$

Remark 2. (Local linear nonparametric estimation) A popular choice in place of the estimator (2.4) in practical work is the local linear estimator (e.g., Fan and Gijbels, 1996)

$$\tilde{g}(x) = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i},$$

where $w_i = K_h(X_i - x) \{S_{n2} - (X_i - x) S_{n1}\}$ and $S_{nj} = \sum_{i=1}^n (X_i - x)^j K_h(X_i - x)$ for $j = 1, 2$. Following the same lines as the proof of Theorem 2.1 and under the same conditions A1-A6 we find that

$$\sqrt{nh/L_n} \left(\tilde{g}(x) - g(x) - \frac{1}{2} h^2 \mu_2(K) g''(x) \right) \rightarrow_d N(0, \sigma^2 \nu_2(K)). \quad (2.7)$$

Thus, $\tilde{g}(x)$ is consistent and has the same limit distribution and bias as the local level estimator (2.5). In the study of nonstationary nonparametric cointegrating regression, Phillips and Wang (2010) found that the local linear and local level estimators also share the same asymptotic distribution and bias. So, local linear regression has no advantage over local level regression in terms of bias reduction in both nonstationary and location shift regressions.

Remark 3. (Bandwidth choice) Defining the limiting bias function as $B(x) = \frac{1}{2} \mu_2(K) g''(x)$, Theorem 2.1 implies that the leading term in the asymptotic mean squared error (MSE) of $\hat{g}(x)$ is given by

$$MSE(\hat{g}(x)) = h^4 B(x)^2 + \frac{L_n \sigma^2 \nu_2(K)}{nh}.$$

Given x , the optimal bandwidth in terms of minimizing $MSE(\hat{g}(x))$ is

$$h^*(x) = \left[\frac{\sigma^2 \nu_2(K)}{4B(x)^2} \right]^{1/5} \frac{1}{(n/L_n)^{1/5}}.$$

To estimate the whole regression function $g(\cdot)$, one can choose a single h to minimize the integrated mean square error $\int MSE(\hat{g}(x)) w(x) dx$, where $w(\cdot)$ is a weight function that

serves to avoid division by zero and to perform trimming in areas of sparse support. Then the optimal bandwidth is given by

$$h^* = \left[\frac{\sigma^2 \nu_2(K) \int w(x) dx}{4 \int B(x)^2 w(x) dx} \right]^{1/5} \frac{1}{(n/L_n)^{1/5}}.$$

The optimal rate $(n/L_n)^{-1/5}$ is larger than the optimal rate $n^{-1/5}$ for the conventional stationary (exogenous) regressor case. Nevertheless, L_n is unobserved in practice, at least without further specification involving measurable variables that determine location shifts. So, in general, we cannot use the plug-in principle to obtain an estimate of h^* or $h^*(x)$. Instead, we can use least squares cross-validation (LSCV) to choose the smoothing parameters. That is, choose h to minimize the following LSCV criterion

$$CV(h) = n^{-1} \sum_{i=1}^n (Y_i - \hat{g}_{-i}(X_i))^2 w(X_i), \quad (2.8)$$

where $\hat{g}_{-i}(X_i)$ is the leave-one-out local constant estimate of $g(X_i)$ and $w(\cdot)$ is some given weight function. We conjecture that the cross-validated bandwidth will converge to the optimal rate $(n/L_n)^{-1/5}$ at some rate, but do not pursue this study here.

Remark 4. (Comparison with nonparametric IV regression) Consider the general nonparametric instrumental variable (IV) regression

$$y_i = g(X_i) + u_i, \quad E(u_i|X_i) \neq 0, \quad E(u_i|W_i) = 0, \quad (2.9)$$

where $(y_i, X_i, W_i)_{i=1}^n$ are observed and W_i is used as an instrument for X_i . Observe that

$$E(y_i|W_i) = E\{g(X_i)|W_i\} = \int g(x) \frac{f_{xw}(x, W_i)}{f_w(W_i)} dx,$$

where $f_{xw}(\cdot, \cdot)$ and $f_w(\cdot)$ are the joint and marginal p.d.f.'s. An estimate of $g(\cdot)$ may be obtained by various functional inversion techniques. However, the inversion of the associated integral operator equation is typically ill-posed because inversion involves an operator that is not bounded or continuous.

In order for W_i to be a valid IV for X_i , we usually require that W_i be observed and that the association between the variables be strong enough for successful inversion (or operator inversion in the functional case) of the estimating equations. In the linear case it is usually sufficient to require that $\text{corr}(X_i, W_i) \neq 0$. To complete the specification of the model (2.9), we add a reduced form equation for the endogenous regressor X_i . Let $m(W_i) = E(X_i|W_i)$ and $u_{xi} = X_i - E(X_i|W_i)$, so that we have

$$X_i = m(W_i) + u_{xi}. \quad (2.10)$$

This reduced form equation helps in identifying the structural curve $g(\cdot)$ in (2.9) provided W_i is observable and the systematic component $m(W_i)$ of X_i provides sufficient leverage in estimation.

To make the link between this model and the location shift system (2.1) - (2.3) explicit, suppose the instrument variable in (2.9) has a triangular array structure so that the systematic component has the form $m(W_{in})$. If the variance of $m(W_{in})$ *expands* as the sample size n increases, then the signal in the regressor X_i correspondingly increases, relative to the variance of the stationary error u_{xi} in (2.10), the variance of the structural equation error u_i in (2.9) and the covariance $\text{cov}(u_{xi}, u_i)$. The possibility of consistent estimation then opens up even in the face of endogeneity in the regressor, a feature that is already well known in the analysis of simple linear parametric models (e.g. Hamilton, 1994, p. 234). As n increases, each distinct value of $m(W_{in})$ may be regarded as carrying location specific information that corresponds to a potentially new location in the continuous location-shift system (2.1) - (2.3). Interestingly in this case, one does not even need to observe W_{in} in order to identify and consistently estimate the true structural regression curve $g(\cdot)$ asymptotically. It is sufficient for this leverage from W_{in} to be present in the regressor X_i . In effect, X_i then has an array structure itself and it is this “nonstationarity” in the regressor that opens up the possibility of consistent estimation by direct nonparametric regression.

Remark 5. (Link with the nonstationary nonparametric cointegrating regression) Recently Wang and Phillips (2009b, 2010) studied structural models of nonparametric cointegration and developed a limit theory for kernel cointegrating regression. It was shown that no ill-posed inverse problem arises in nonparametric models with nonstationary endogenous regressors and identification does not require the existence of observable instrumental variables that are orthogonal to the structural equation errors. The location-shift model shares these features in much the same spirit. To see the analogy between the two types of models, consider the following nonlinear structural model of cointegration

$$y_i = g(X_i) + u_i, \quad i = 1, 2, \dots, n,$$

where u_i is a zero mean stationary process, X_i is a nonstationary $I(1)$ regressor, and $g(\cdot)$ is an unknown smooth function to be estimated. Wang and Phillips (2009b, 2010) show that under some regularity conditions, the local constant estimate $\hat{g}(x)$ of $g(x)$ has the following limiting mixed normal distribution:

$$(nh^2)^{1/4} \left(\hat{g}(x) - g(x) - \frac{h^2 \mu_2(K)}{2} g''(x) \right) \rightarrow_d \frac{\sigma_u \mathcal{N}}{L^{1/2}(1, 0)}$$

where σ_u is a constant that depends on the kernel and the parameters underlying the process $\{X_i, u_i\}$, and \mathcal{N} is a standard normal variate independent of the local time process $\{L(t, 0)\}$ of the Brownian motion associated with the limit of the standardized process $n^{-1/2}X_{[n\cdot]}$. The kernel estimates in the current paper are similar to the nonparametric cointegrating estimates of

Wang and Phillips (2009b) in at least four aspects. First, both estimates deal with endogeneity in the regressor without using instrumental variables. Instead, identification occurs through the expansion of the variance of the regressors: either continuous location shifts or unit-root behavior ensures that $\text{Var}(X_i)$ expands as the sample size increases and that data accumulates steadily over the entire support. In effect, continuous location shifts provide a form of recurrence in X_i , corresponding to the capacity of a nonstationary random wandering process to visit all points in the space an infinite number of times. This type of recurrent behavior in X_i ensures that the structural regression curve is effectively traced out continuously by the data and is correspondingly amenable to nonparametric regression, just as Fig. 2 suggests. Second, as mentioned in Remark 3, both estimates have the same asymptotic distribution and bias as the corresponding local linear estimates. Third, both estimates have a slower convergence rate than nonparametric regression with a stationary regressor, namely, $(nh)^{1/2}$. As noticed by Wang and Phillips (2009a), in the unit-root case, the amount of time spent by the process around any particular spatial point is of order \sqrt{n} rather than n so that the corresponding convergence rate in nonparametric cointegrating regression is $\sqrt{\sqrt{nh}} = (nh^2)^{1/4}$. In the case of continuous location shifts, the number of effective observations at each location is of order nh/L_n which gives the convergence rate $\sqrt{nh/L_n}$. Fourth, for both types of estimates, the conditional variance $\sigma^2(x) = E(u_i^2|X_i = x)$ does not play a role in the asymptotics. Instead, it is the unconditional variance that really matters. This result indicates that the width of the pointwise confidence bands for either type of estimate should remain largely the same across evaluation points in areas with abundant data.

2.2 Models with compact support

The formulation of the location-generating mechanism in (2.2) - (2.3) requires that the regressor locations eventually cover the whole real line as the sample size grows in order to accommodate the infinite support of the random elements in the structural model. This condition facilitates identification and the development of the asymptotic theory, but it also creates two problems. First, the usual convergence rate of the nonparametric local constant and local linear estimates is reduced in the case of continuous location shifts in an endogenous regressor because some portion of the data is used up to achieve identification by tracing out the curve through shifts in location. Second, the sample variance of the endogenous regressor needs to *expand* as the sample size n increases. A natural question is whether these two conditions are vital to identification and consistent nonparametric estimation.

It turns out the expansion of regressor location over the whole real line is unnecessary when u_{xi} is compactly supported. As indicated earlier, the nonparametric IV literature frequently assumes compact support for the endogenous regressor in order to use standard functional analysis. We therefore consider a similar case where the location shifts occur in a compact set. The model is as specified in (2.1) - (2.3), but we now assume that the $2m + 1$ locations

are spaced over a fixed compact interval $[-L/2, L/2]$, i.e., $L_n = L$ does not pass to infinity as $n \rightarrow \infty$. As before, since we do not restrict u_{xi} to have zero mean, it is not restrictive to require that the $2m + 1$ locations are spaced over a compact interval that is symmetric around 0. We continue to assume that $M = \sum_{i=1}^n 1\{i \in A_\alpha\}$ so that $n = (2m + 1)M$. The pair (u_i, u_{xi}) continues to satisfy Assumptions A1, A2, and A3(i).

To proceed, we make the following assumptions.

Assumption A7. *The error term u_{xi} has compact support, i.e., $u_{xi} \in [\underline{u}, \bar{u}]$ a.s. for some finite numbers \underline{u} and \bar{u} .*

Assumption A8. *For given x , L is sufficiently large such that $x \in (-L/2 + \bar{u}, L/2 + \underline{u})$.*

Assumption A9. *As $n \rightarrow \infty$, $m \rightarrow \infty$, $h \rightarrow 0$, $nh \rightarrow \infty$, $nh^5 \rightarrow c \in [0, \infty)$, and $Mh/m^3 \rightarrow 0$.*

Note that under A7, the tail conditions in Assumptions A3(ii)-(iii) are redundant. A8 requires that $L > \bar{u} - \underline{u}$. Intuitively, it implies that the larger L is, the greater the portion of the true regression curve that can be identified and consistently estimated. A9 parallels A6.

The following theorem establishes the consistency and asymptotic normality of $\hat{g}(x)$.

Theorem 2.2 *Suppose Assumptions A1-A2, A3(i), A4, A5 and A7-A9 hold. Then*

$$\begin{aligned} \sqrt{nh/L} \left(\hat{g}(x) - g(x) - h^2 \mu_2(K) \left\{ g'(x) \int_{\underline{u}}^{\bar{u}} f'_{u_x}(p) dp + \frac{1}{2} g''(x) \right\} \right) \\ \rightarrow_d N(0, \sigma^2 \nu_2(K)). \end{aligned} \quad (2.11)$$

Theorem 2.2 indicates that the local constant estimate in the case of continuous location shifts can achieve the usual \sqrt{nh} -rate of consistency. The same is also true for the local linear estimate. This fast rate contrasts with the much slower rates achievable for nonparametric IV estimation without the advantage of location shifts. In comparison with the result in Theorem 2.1, we observe that the bias function in (2.11) contains the linear term, $g'(x) \int_{\underline{u}}^{\bar{u}} f'_{u_x}(p) dp$, which vanishes if $f_{u_x}(\underline{u}) = f_{u_x}(\bar{u})$, as in the infinite support case. If the last condition holds, then the local constant and local linear estimates share the same asymptotic distribution and bias as well. Otherwise, they only share the same asymptotic distribution after bias correction. The following section reveals a further advantage of local smoothing techniques in dealing with issues of endogeneity.

3 Inconsistent Parametric Regression under Location Shifts

If the regression function is parametric and if its functional form is known, parametric estimation becomes possible. In particular, if $g(\cdot)$ in (2.1) is known to be of the parametric form $g(x) = g(x, \theta)$, say, where $g(\cdot, \cdot)$ is known up to the finite dimensional parameter, θ , then $g(\cdot)$

can be estimated by direct parametric estimation of θ . However, under the conditions given in Theorem 2.2 and assuming that the form of $g(x) = g(x, \theta)$ is known, the nonlinear least squares (NLS) estimate $\hat{\theta}$ of θ is generally inconsistent because of endogeneity in the regressor. To see this, we focus on the simple linear model

$$y_i = \beta_0 + \beta_1 X_i + u_i, \quad E(u_i | X_i) \neq 0, \quad E(u_i) = 0, \quad (3.1)$$

where the endogenous regressor X_i is generated according to (2.2)-(2.3). Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the ordinary least squares (OLS) estimators of β_0 and β_1 , respectively. Below we first demonstrate the inconsistency of these OLS estimators and then propose two methods for consistent estimation of (β_0, β_1) .

3.1 Inconsistency of $(\hat{\beta}_0, \hat{\beta}_1)$

To study the asymptotic properties of the OLS estimator $(\hat{\beta}_0, \hat{\beta}_1)$, we maintain Assumptions A1-A2 and add the following assumption.

Assumption A2*. $E(u_{xi}) = \mu_x$, and $\text{Var}(u_{xi}) = \sigma_x^2 < \infty$.

Note that the above assumption is automatically satisfied under Assumption A7 for the case of regressors with compact support. Under Assumptions A1, A2 and A2*, we first derive the probability limit of the OLS estimator $\hat{\beta}_1$ and show that it is inconsistent for β_1 when L_n is fixed. The probability limit of $\hat{\beta}_0$ follows straightforwardly. Write

$$\hat{\beta}_1 = \beta_1 + \frac{n^{-1} \sum_{i=1}^n (X_i - \bar{X}) u_i}{n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2}, \quad (3.2)$$

where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$. First, by the definition of $\{\mu_\alpha\}$ and the weak law of large numbers (WLLN), we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 &= \frac{1}{2m+1} \sum_{\alpha=-m}^m \frac{1}{M} \sum_{i \in A_\alpha} \left[\frac{L_n \alpha}{2m} + (u_{xi} - \bar{u}_x) \right]^2 \\ &= \frac{L_n^2}{4m^2(2m+1)} \sum_{\alpha=-m}^m \alpha^2 + \frac{1}{n} \sum_{i=1}^n (u_{xi} - \bar{u}_x)^2 + \frac{L_n}{m(2m+1)} \sum_{\alpha=-m}^m \frac{\alpha}{M} \sum_{i \in A_\alpha} (u_{xi} - \bar{u}_x) \\ &= \frac{L_n^2}{2m^2(2m+1)} \sum_{\alpha=1}^m \alpha^2 + \sigma_x^2 + o_P(1) \\ &= \frac{L_n^2}{12} \{1 + o(1)\} + \sigma_x^2 + o_P(1), \end{aligned} \quad (3.3)$$

where $\bar{u}_x = n^{-1} \sum_{i=1}^n u_{xi}$, and the third line follows from the WLLN and Chebyshev inequality provided $n/L_n^2 \rightarrow \infty$ as $n \rightarrow \infty$.¹

¹Letting $T_n \equiv \frac{L_n}{m(2m+1)} \sum_{\alpha=-m}^m \frac{\alpha}{M} \sum_{i \in A_\alpha} (u_{xi} - \mu_x) = \frac{L_n}{m(2m+1)} \sum_{\alpha=-m}^m \frac{\alpha}{M} \sum_{i \in A_\alpha} (u_{xi} - \bar{u}_x)$, then $E(T_n) = 0$, and $\text{Var}(T_n) = \frac{L_n^2}{Mm^2(2m+1)^2} \sum_{\alpha=-m}^m \alpha^2 \sigma_x^2 = O\left(\frac{L_n^2}{Mm}\right) = O\left(\frac{L_n^2}{n}\right) = o(1)$.

Next

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) u_i &= \frac{1}{2m+1} \sum_{\alpha=-m}^m \frac{1}{M} \sum_{i \in A_\alpha} \left(\frac{L_n \alpha}{2m} + u_{xi} - \bar{u}_x \right) u_i \\
&= \frac{L_n}{2m(2m+1)} \sum_{\alpha=-m}^m \frac{1}{M} \sum_{i \in A_\alpha} \alpha u_i + \frac{1}{n} \sum_{i=1}^n u_{xi} u_i - \frac{\bar{u}_x}{n} \sum_{i=1}^n u_i \\
&= \frac{L_n}{2m(2m+1)} \sum_{\alpha=-m}^m \frac{1}{M} \sum_{i \in A_\alpha} \alpha u_i + E(u_{xi} u_i) + o_P(1).
\end{aligned}$$

Consider the first term in the last expression. Let $T_{1n} = \frac{L_n}{2m(2m+1)} \sum_{\alpha=-m}^m \frac{1}{M} \sum_{i \in A_\alpha} \alpha u_i$. Then $E(T_{1n}) = 0$ as $E(u_i) = 0$, and

$$\begin{aligned}
\text{Var}(T_{1n}) &= \frac{L_n^2}{4m^2(2m+1)^2} \text{Var} \left(\sum_{\alpha=-m}^m \frac{1}{M} \sum_{i \in A_\alpha} \alpha u_i \right) \\
&= \frac{L_n^2 \sigma^2}{4Mm^2(2m+1)^2} \sum_{\alpha=-m}^m \alpha^2 = O \left(\frac{L_n^2}{n} \right) = o(1).
\end{aligned}$$

Hence $T_{1n} = o_P(1)$ and $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) u_i = E(u_{xi} u_i) + o_P(1)$. This, together with (3.2) and (3.3) and, implies

$$\hat{\beta}_1 = \beta_1 + \frac{E(u_{xi} u_i) + o_P(1)}{\frac{L_n^2}{12} \{1 + o(1)\} + \sigma_x^2 + o_P(1)} = \beta_1 + \frac{E(u_{xi} u_i)}{\frac{L_n^2}{12} + \sigma_x^2} \{1 + o_P(1)\}, \quad (3.4)$$

thereby giving the following result.

Lemma 3.1 *Suppose Assumptions A1, A2, and A2* hold and $n/L_n^2 \rightarrow \infty$. Then*

$$\hat{\beta}_1 = \beta_1 + \frac{E(u_{xi} u_i)}{\frac{L_n^2}{12} + \sigma_x^2} \{1 + o_P(1)\}.$$

The following remark explores the implications of the above lemma, considering the two cases where $L_n = L$ is fixed and $L_n \rightarrow \infty$ as $n \rightarrow \infty$.

Remark 6. (Inconsistency of parametric regression) In the first case ($L_n = L$ fixed), $\hat{\beta}_1$ has the probability limit

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_1 = \beta_1 + \frac{E(u_{xi} u_i)}{\frac{L^2}{12} + \sigma_x^2},$$

and is inconsistent unless $E(u_{xi} u_i) = 0$, viz., X_i is exogenous. For $\hat{\beta}_0$, we have

$$\hat{\beta}_0 - \beta_0 = -\bar{X} (\hat{\beta}_1 - \beta_1) + n^{-1} \sum_{i=1}^n u_i = -\mu_x \frac{E(u_{xi} u_i)}{\frac{L^2}{12} + \sigma_x^2} + o_P(1). \quad (3.5)$$

so $\hat{\beta}_0$ is inconsistent for β_0 unless either $\mu_x = 0$ or $E(u_{xi} u_i) = 0$. Hence, the parametric estimator $\hat{g}^p(x) = \hat{\beta}_0 + \hat{\beta}_1 x$ is inconsistent for $g(x) = \beta_0 + \beta_1 x$ at all points except $x = \mu_x$.

By contrast, according to Theorem 2.2, the nonparametric estimator is consistent for all x satisfying certain domain restrictions.

In the second case ($L_n \rightarrow \infty$), the OLS estimator $\hat{\beta}_1$ is consistent for β_1 as $\hat{\beta}_1 = \beta_1 + o_P(1)$, due to the strengthening signal in the regressor as $L_n \rightarrow \infty$. This result, together with (3.5), implies that $\hat{\beta}_0$ is consistent for β_0 , and so the parametric regression estimator of $g(x) = \beta_0 + \beta_1 x$ is also consistent. However, if L_n diverges to infinity slowly like $L_n = \log n$, the estimation bias may disappear at a very slow rate.

To find the limit distribution of $(\hat{\beta}_0, \hat{\beta}_1)$, we add the following assumption.

Assumption 2.** $E[|u_i u_{xi}|^{2+\delta}] < \infty$ for some $\delta > 0$.

Let $\theta = (\beta_0, \beta_1)'$ and $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1)'$. Let $\underline{X}_i = (1, X_i)'$, $\mathbf{X} = (\underline{X}_1, \dots, \underline{X}_n)'$, $\mathbf{u} = (u_1, \dots, u_n)'$, and $\mathbf{y} = (y_1, \dots, y_n)'$. Define $D_n = \text{diag}(1, L_n)$,

$$\Gamma = \lim_{n \rightarrow \infty} \begin{pmatrix} 1 & \frac{\mu_x}{L_n} \\ \frac{\mu_x}{L_n} & \frac{1}{12} + \frac{E(u_{xi}^2)}{L_n^2} \end{pmatrix}, \text{ and } \Omega = \lim_{n \rightarrow \infty} \begin{pmatrix} \sigma^2 & \frac{E(u_{xi}u_i^2)}{L_n} \\ \frac{E(u_{xi}u_i^2)}{L_n} & \frac{\sigma^2}{12} + \frac{\text{Var}(u_{xi}u_i)}{L_n^2} \end{pmatrix}.$$

After centering, the limiting distribution of $\hat{\theta}$ is given in the following theorem.

Theorem 3.2 Suppose Assumptions A1, A2, A2*, and A2** hold and $n/L_n^2 \rightarrow \infty$. Then

$$\sqrt{n}D_n \left(\hat{\theta} - \theta - (\mathbf{X}'\mathbf{X})^{-1} E(\mathbf{X}'\mathbf{u}) \right) \rightarrow_d N(0, \Gamma^{-1}\Omega\Gamma^{-1}). \quad (3.6)$$

Remark 7. (Different convergence rates for the intercept and slope estimates)

Straightforward calculations show that

$$(\mathbf{X}'\mathbf{X})^{-1} E(\mathbf{X}\mathbf{u}) = \frac{E(u_{xi}u_i)}{n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2} \begin{pmatrix} -\bar{X} \\ 1 \end{pmatrix}.$$

An immediate implication of Theorem 3.2 is therefore that

$$\sqrt{n} \left(\hat{\beta}_0 - \beta_0 + \frac{\bar{X} E(u_{x1}u_1)}{n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2} \right) \rightarrow_d N(0, \omega_{11}), \quad (3.7)$$

$$L_n \sqrt{n} \left(\hat{\beta}_1 - \beta_1 - \frac{E(u_{x1}u_1)}{n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2} \right) \rightarrow_d N(0, \omega_{22}), \quad (3.8)$$

where ω_{11} and ω_{22} are the (1, 1) and (2, 2) elements of $\Gamma^{-1}\Omega\Gamma^{-1}$, respectively. If $L_n \rightarrow \infty$ as $n \rightarrow \infty$, then

$$\Gamma^{-1}\Omega\Gamma^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{12} \end{pmatrix}^{-1} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{12} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{12} \end{pmatrix}^{-1} = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 12 \end{pmatrix}.$$

In this case, the intercept estimator, after centering, is asymptotically independent of the slope estimator, and the slope coefficient estimator has a faster convergence rate, due to the stronger

signal in the regressor. In contrast, in the fixed L_n case, the two estimators are asymptotically dependent, have the same rate of convergence, and the range of location shifts contributes to the asymptotic variance formula in (3.6) in a complicated way.

In spite of the $O(\sqrt{n})$ convergence rate for the intercept estimator and the $O(L_n\sqrt{n})$ convergence rate for the slope estimator, the result in (3.6) does not seem useful for inferential purposes because the bias term in (3.6) does not appear to be estimable at the required \sqrt{n}/L_n^2 -rate (for the intercept parameter) or \sqrt{n}/L_n -rate (for the slope parameter) to be eliminated (recall from (3.3) that $n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2 = O_P(L_n^2)$). It is worth mentioning that the residuals $\{\hat{u}_i\}$ from the OLS regression are useless in constructing a consistent estimate of $E(X_i u_i)$ because of the orthogonality of \hat{u}_i and X_i . One may consider consistent estimation of $E(X_i u_i)$ by estimating the model first via the local level (or local linear) nonparametric method at interior points and then obtaining the nonparametric residuals from the structural equation. Unfortunately, this approach usually requires uniform consistency of the local level (or local linear) estimate over the whole support of the regressor, which seems extremely difficult here given the fact that the variance of the regressor is expanding as the sample size increases or that the nonparametric estimates are only consistent at points within a subset of the interior of the support of the regressor for the fixed L_n case. Instead, the next subsection proposes two alternative methods to achieve consistent estimation of (β_0, β_1) by direct use of the nonparametric level estimate in a parametric regression.

3.2 Spatial \mathcal{L}_2 and Bias-Corrected OLS Estimation

In this subsection we propose two methods for consistent estimation of (β_0, β_1) . The first method is a spatial \mathcal{L}_2 regression and the second method involves bias-corrected OLS estimation. The \mathcal{L}_2 method treats the linear regression function as unknown, estimates it nonparametrically by $\hat{g}(x)$, and then regresses $\hat{g}(x)$ on $(1, x)$ to estimate the unknown parameter (β_0, β_1) by minimizing a spatial \mathcal{L}_2 criterion function, using a continuum of pseudo-observations on $(x, \hat{g}(x))$ where x is restricted to be bounded away from the two tails. We prove that the resulting \mathcal{L}_2 estimator of β_0 can be \sqrt{n} -consistent and that of β_1 can be super-consistent. We show that the OLS bias terms can be corrected by using the residuals from the \mathcal{L}_2 regression, and the bias-corrected OLS estimators can achieve the same consistency rates as the \mathcal{L}_2 estimators.

3.2.1 Spatial \mathcal{L}_2 regression

Noting that the local level estimate $\hat{g}(x)$ is consistent for $g(x) = \beta_0 + \beta_1 x$ in the interior of the regressor support, we propose to estimate the unknown parameter $\theta \equiv (\beta_0, \beta_1)'$ by minimizing

the following (spatial) \mathcal{L}_2 criterion

$$S_n(\beta_0, \beta_1) = \int_{a_n}^{b_n} (\hat{g}(x) - \beta_0 - \beta_1 x)^2 \hat{f}(x) dx \quad (3.9)$$

where a_n and b_n are integration limits that serve to truncate observations in the two tails, $\hat{f}(x) = N^{-1} \sum_{i=1}^n K_h(x - X_i)$ is a pseudo-estimate of the “density” of X_i , and $N = n/L_n$ signifies the effective number of observations used in the nonparametric estimation, which does not need to be observed in practice for implementation. In the case where $L_n \rightarrow \infty$ as $n \rightarrow \infty$, we allow but do not require that the integration limits a_n and b_n pass to (positive or negative) infinity: if they pass to infinity, we assume for simplicity that they do so at the same rate c_n and write $a_n = c_n a$ and $b_n = c_n b$, where $c_n \rightarrow \infty$ and $c_n = O(L_n)$ due to the endogeneity bias of the nonparametric estimators in the tails; if they do not pass to infinity, we can simply write a_n and b_n as a and b by taking $c_n = 1$. In the case where $L_n = L$ is fixed as $n \rightarrow \infty$, we can only consider fixed integration limits and take $c_n = 1$. Note that $\hat{f}(x)$ is a weight function that serves to avoid division by zero and to perform trimming in areas of sparse support, and $[a_n, b_n]$ defines a (possibly expanding) compact set on which the nonparametric estimates $\hat{g}(x)$ are used in the estimation of (β_0, β_1) . Clearly, (3.9) provides a trimming operation implicitly via the local nature of the estimate $\hat{g}(x)$ and explicitly via the use of the (truncated) domain $[a_n, b_n]$.

The minimizer of (3.9) $\tilde{\theta} \equiv (\tilde{\beta}_0, \tilde{\beta}_1)'$ is given by

$$\tilde{\theta} = Q_n^{-1} \begin{pmatrix} \int_{a_n}^{b_n} \hat{g}(x) \hat{f}(x) dx \\ \int_{a_n}^{b_n} x \hat{g}(x) \hat{f}(x) dx \end{pmatrix} \text{ where } Q_n = \begin{pmatrix} \int_{a_n}^{b_n} \hat{f}(x) dx & \int_{a_n}^{b_n} x \hat{f}(x) dx \\ \int_{a_n}^{b_n} x \hat{f}(x) dx & \int_{a_n}^{b_n} x^2 \hat{f}(x) dx \end{pmatrix}.$$

To develop the limit theory, we make the following assumptions.

Assumption A3*. *Either one of the following conditions holds:*

(i) *The error term u_{xi} has infinite support such that Assumption A.3(ii) holds, $L_n^{-v} = O(h^2)$, and $f_{u_x}(L_n) = O(h^2)$ as $L_n \rightarrow \infty$. As $n \rightarrow \infty$, $L_n \rightarrow \infty$.*

(ii) *The error term u_{xi} has compact support, i.e., $u_{xi} \in [\underline{u}, \bar{u}]$ a.s. for some finite numbers \underline{u} and \bar{u} . L_n is either fixed or tends to ∞ as $n \rightarrow \infty$. If $L_n = L$ is fixed, L is sufficiently large that $x \in (-L/2 + \bar{u}, L/2 + \underline{u})$ for all $x \in [a, b]$.*

Assumption A6*. *As $n \rightarrow \infty$, $n/L_n^2 \rightarrow \infty$, $nh^4 c_n/L_n \rightarrow 0$, $nm^{-4} c_n L_n^3 \rightarrow 0$, $(n/L_n)^{-\delta/2} L_n \rightarrow 0$, and c_n either passes to infinity or takes value 1 such that $c_n = O(L_n)$.*

Assumptions A3* is modified from Assumption A3(ii)-(iii) for the infinite support case and A7-A8 for the compact support case. Assumption A6* is adapted from Assumption A6 by taking into account the known linear regression function and imposing an undersmoothing bandwidth sequence ($nh^4 c_n/L_n \rightarrow 0$). The next to last requirement in Assumption A6* is needed to verify the Liapounov condition.

Theorem 3.3 Suppose Assumptions A1, A2, A2*, A2**, A3(i), A3*, A5, and A6* hold. Then

$$\sqrt{n/L_n} C_n (\tilde{\theta} - \theta) \rightarrow_d N(0, \sigma^2 Q^{-1}), \quad (3.10)$$

$$\text{where } C_n = \begin{pmatrix} \sqrt{c_n} & 0 \\ 0 & \sqrt{c_n^3} \end{pmatrix} \text{ and } Q = \begin{pmatrix} b-a & \frac{b^2-a^2}{2} \\ \frac{b^2-a^2}{2} & \frac{b^3-a^3}{3} \end{pmatrix}.$$

Remark 8. (Consistency of $\tilde{\beta}_0$ and $\tilde{\beta}_1$) Despite the nonparametric convergence rate of the regressand $\hat{g}(x)$, Theorem 3.3 indicates that the \mathcal{L}_2 estimate $\tilde{\beta}_0$ is \sqrt{n} -consistent whereas the \mathcal{L}_2 estimate $\tilde{\beta}_1$ is $\sqrt{nL_n}$ -consistent provided one takes $c_n = L_n$. Both estimates achieve the parametric \sqrt{n} -rate of consistency for the case of fixed L_n and $\tilde{\beta}_1$ is super-consistent in the case where L_n passes to infinity as $n \rightarrow \infty$. To obtain these results in Theorem 3.3, we have employed two devices. First, undersmoothing is required to eliminate the $O(h^2)$ bias terms from the first stage nonparametric regression estimate of $g(x)$. This is standard in the nonparametric or semiparametric literature when the first stage nonparametric estimates are used in a second stage parametric or nonparametric estimation. Second, to reduce the variation of the nonparametric estimates, we have used integration in the \mathcal{L}_2 regression. The smoothing operation of integration helps to produce the (nearly) parametric convergence rate of $\tilde{\theta}$ despite the slow nonparametric convergence rate of $\hat{g}(x)$. The mechanism is analogous to that of average marginal effect or derivative estimation.

It is worth mentioning that Cristóbal, Roca, and Manteiga (1987) defined a class of linear regression estimators by minimizing the following \mathcal{L}_2 criterion that is similar to our spatial regression criterion:

$$\overline{S}_n(\beta_0, \beta_1) = \int (g_n(x) - \beta_0 - \beta_1 x)^2 d\Omega_n(x)$$

where $g_n(x)$ is a nonparametric estimate of $g(x) = E(y_i | X_i = x)$ and Ω_n is a weight measure that typically has an infinite support. They demonstrate that this class of estimators includes the ordinary and generalized ridge regression estimators as special cases and they provide some asymptotic properties of these estimators for the classical *iid* design. In contrast, our model has endogenous regressors that involve a complicated location shift pattern, and we intentionally use a compactly supported weight function ($\hat{f}(x) 1\{a_n \leq x \leq b_n\}$), rather than infinite support, in order to provide protection against the bias effects of endogeneity in the tails.

3.2.2 Bias-corrected OLS estimation

We now propose a bias-correction procedure for the OLS estimator of $\theta = (\beta_0, \beta_1)'$ in the linear structural equation (3.1). Let $\tilde{u}_i = y_i - \tilde{\beta}_0 - \tilde{\beta}_1 X_i$ and $\tilde{c} = n^{-1} \sum_{i=1}^n \tilde{u}_i X_i$. We define bias-corrected OLS estimators of β_0 and β_1 , respectively, as

$$\hat{\beta}_{0c} = \hat{\beta}_0 + \frac{\overline{X} \tilde{c}}{n^{-1} \sum_{i=1}^n (X_i - \overline{X})^2} \text{ and } \hat{\beta}_{1c} = \hat{\beta}_1 - \frac{\tilde{c}}{n^{-1} \sum_{i=1}^n (X_i - \overline{X})^2}.$$

Let $\widehat{\theta}_c \equiv (\widehat{\beta}_{0c}, \widehat{\beta}_{1c})'$. The following theorem establishes the consistency and asymptotic normality of $\widehat{\theta}_c$.

Theorem 3.4 *Suppose Assumptions A1, A2, A2*, A2**, A3(i), A5, and A6*. (i) If Assumption A3*(i) holds or A3*(ii) holds with $L_n \rightarrow \infty$ as $n \rightarrow \infty$, then $\sqrt{n/L_n}C_n(\widehat{\theta}_c - \theta) \rightarrow_d N(0, \Psi)$, where*

$$\Psi = \sigma^2 \lim_{n \rightarrow \infty} \begin{pmatrix} c_n^{-2} \mu_x^2 q^{22} + \frac{c_n}{L_n} & -c_n^{-1} \mu_x q^{22} + \frac{c_n}{L_n} \gamma \\ -c_n^{-1} \mu_x q^{22} + \frac{c_n}{L_n} \gamma & q^{22} \end{pmatrix}$$

$\gamma = q^{21}(b-a) + \frac{1}{2}q^{22}(b^2 - a^2)$, and q^{ij} is the (i, j) element of Q^{-1} for $i, j = 1, 2$.

(ii) If Assumption A3*(ii) holds with $L_n = L$ fixed, then $\sqrt{n/L}(\widehat{\theta}_c - \theta) \rightarrow_d N(0, \Psi)$, where

$$\Psi = B^{-1} \Upsilon B^{-1}, \quad \Upsilon = \sigma^2 \begin{pmatrix} L^{-1} & L^{-1} \\ L^{-1} & c'Q^{-1}c \end{pmatrix}, \quad B = \begin{pmatrix} 1 & \mu_x \\ \mu_x & \frac{L^2}{12} + E(u_{xi}^2) \end{pmatrix},$$

and $c = (\mu_x, c_x)'$ with $c_x = \frac{L^2}{12} + E(u_{xi}^2)$.

Remark 9. (Consistency and extension) A similar remark to that after Theorem 3.3 holds for $\widehat{\beta}_{0c}$ and $\widehat{\beta}_{1c}$ in part (i) of the above theorem. In either of the two most important special cases, $c_n = L_n$ and $c_n = 1$, the formula for Ψ can be greatly simplified. Even though we only focus on the linear structural equation model as specified in (3.1), it is straightforward to extend our theory to the general nonlinear structural equation model. Suppose $\{y_i\}$ is generated according to

$$y_i = g(X_i, \theta) + u_i, \quad E(u_i | X_i) \neq 0, \quad (3.11)$$

where $g(\cdot, \theta)$ is known up to the finite dimensional parameter θ , and the endogenous regressor X_i satisfies (2.2) and (2.3). It is straightforward to show that the NLS estimator $\widehat{\theta}$ of θ is inconsistent. As before, the nonparametric local level estimate $\widehat{g}(x)$ of $g(x) = g(x, \theta)$ is consistent for a large portion of the domain of the regressor. Then, the unknown parameter θ can be estimated by minimizing the following (spatial) \mathcal{L}_2 criterion

$$S_n(\theta) = \int_{a_n}^{b_n} (\widehat{g}(x) - g(x, \theta))^2 \widehat{f}(x) dx, \quad (3.12)$$

just as before. Let $\widetilde{\theta}$ denote the solution to the above minimization problem. Following the proof of Theorem 3.3, we can establish the consistency of $\widetilde{\theta}$ under some regularity conditions. A bias-corrected NLS estimator can also be constructed. The details are similar to the linear case and are omitted.

4 Concluding remarks

This paper shows that location shifts in a regressor can play an effective role in tracing out a regression curve in spite of endogeneity in the regressor. In part, these location shifts act as an instrument that moves the data along the curve, and in part they add variation to the regressor that enhances the signal/noise ratio. In both respects, such location shifts act in a manner analogous to the random wandering feature of unit root regressors in a cointegrating regression equation, thereby explaining the consistency of simple nonparametric regression in both cases.

Importantly, in all these cases there is no need for nonparametric IV estimation or the complications of functional inversion and regularization. Just as in textbook discussions of identification, regression curves may be identified and estimated by nonparametric methods provided the data embody a mechanism for tracing out the regression curve of interest. Our results also reveal the significance of the common assumption in the nonparametric IV literature of a compact support. In such cases, nonparametric regression can have an important advantage over parametric regression (even when the true form of the regression function is known) in terms of its consistency and a \sqrt{nh} convergence rate in contrast to the inconsistency of parametric regression.

As our main results show, location shifts remove the effects of endogeneity and ensure that the local linear and local constant estimates have the same asymptotic distribution and bias in the case of infinite support. This result is analogous to nonparametric cointegrating regression where the limit theory also involves only a single $O(h^2)$ bias term (Wang and Phillips, 2009b). Interestingly, simulations (not reported here) reveal that when there is strong endogeneity in the regression, local linear estimation is outperformed by local level regression near the boundary and appears to be more susceptible to endogeneity bias in this region of the sample space.

We also explore a paradox where the greater use of correct prior information on a model can be detrimental in regression. The explanation for this paradox is that even when we use additional correct information about the specification of a model, that information may still not be complete and, in consequence, may distort regression results. In the example studied here, the correct additional information used is considerable and is the full functional form specification of the model. Nevertheless, the omitted information (endogeneity) that makes the model specification incomplete is very important and leads to inconsistency in parametric regression.

In such situations, it is very interesting that partial information can be successful where more complete information fails. In applied statistics, it has long been known that controlling for outliers in regression can help to achieve robustness. We show that nonparametric kernel regression naturally utilizes this mechanism to great advantage in structural regression. More specifically, kernel regression has a considerable additional advantage beyond its usually touted advantage of robustness to (unknown) functional form. Local nonparametric regression also

provides robustness to endogeneity in the regressor when there are systematic influences that assure identification, such as location shifts or nonstationarity in the data.

It is also possible to obtain consistent estimation by parametric methods in such cases. In particular, spatial \mathcal{L}_2 regression is shown to successfully remove endogeneity bias and inconsistency by bounding the domain of the regression. This approach is analogous to the treatment of outliers – it provides protection against possible effects of endogeneity in parametric regression by paying a premium through the loss of tail information in the data.

5 Appendix: technical results and proofs

5.1 Proof of Theorem 2.1

We first state and prove two lemmas that are used in the proof of Theorem 2.1.

Lemma 5.1 *Let $f_u(\cdot)$ be the p.d.f. of $\{u_i\}$. Suppose A2-A3 hold. Then*

$$\int u \left[L_n (2m)^{-1} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha) - f_u(u) \right] du = O(L_n^2/m^2 + L_n^{-v}).$$

Proof. The trapezoidal rule approximates $\int_a^b g(p) dp$ by $\frac{1}{2}(b-a)(g(a) + g(b))$ and the approximation error is given by the Newton-Cotes formula (e.g., Stoer and Bulirsch (1993, p. 162))

$$\frac{1}{2}(b-a)(g(a) + g(b)) - \int_a^b g(p) dp = \frac{1}{12}(b-a)^3 g''(c), \quad (5.1)$$

where the second order derivative $g''(\cdot)$ of $g(\cdot)$ is continuous and $c \in (a, b)$. Let $\beta_n = 2m/L_n$. Write

$$\begin{aligned} & \frac{L_n}{2m} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha) - f_u(u) \\ &= \left\{ \frac{1}{\beta_n} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha) - \int_{-L_n/2}^{L_n/2} f(u, x - p) dp \right\} \\ & \quad + \left\{ \int_{-L_n/2}^{L_n/2} f(u, x - p) dp - \int_{-\infty}^{\infty} f(u, x - p) dp \right\} \\ &= \left\{ \sum_{\alpha=-m}^m \frac{1}{2} \left[f\left(u, x - \frac{\alpha-1}{\beta_n}\right) + f\left(u, x - \frac{\alpha}{\beta_n}\right) \right] \frac{1}{\beta_n} - \int_{-L_n/2}^{L_n/2} f(u, x - p) dp \right\} \\ & \quad + \left\{ \int_{-L_n/2}^{L_n/2} f(u, x - p) dp - \int_{-\infty}^{\infty} f(u, x - p) dp \right\} \\ & \quad + \frac{1}{2\beta_n} \left\{ f\left(u, x - \frac{L_n}{2}\right) - f\left(u, x + \frac{(m+1)L_n}{2m}\right) \right\} \\ &\equiv I_{1m}(u) + I_{2m}(u) + I_{3m}(u). \end{aligned}$$

By (5.1) and Assumption A3(i)

$$\begin{aligned}\beta_n^2 \int_{-\infty}^{\infty} u I_{1m}(u) du &= \int_{-\infty}^{\infty} \frac{u}{12\beta_n} \sum_{\alpha=-m}^m f_2''(u, x - p_\alpha) du \\ &\rightarrow \frac{1}{12} \int_{-\infty}^{\infty} u \int_{-\infty}^{\infty} f_2''(u, x - p) dp du = \frac{1}{12} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u f_2''(u, p) dp du,\end{aligned}$$

where p_α lies between $(\alpha - 1)/\beta_n$ and α/β_n . By Assumptions A2 and A3 (ii)-(iii)

$$\begin{aligned}\left| \int_{-\infty}^{\infty} u I_{2m}(u) du \right| &\leq \int_{-\infty}^{\infty} |u| \left| \int_{-L_n/2}^{L_n/2} f(u, x - p) dp - \int_{-\infty}^{\infty} f(u, x - p) dp \right| du \\ &\leq (L_n/2)^{-v} \int_{-\infty}^{\infty} |u| C_f(u) du = O(L_n^{-v}),\end{aligned}$$

and

$$\begin{aligned}\left| \int_{-\infty}^{\infty} u I_{3m}(u) du \right| &\leq \frac{1}{2\beta_n} \left\{ \int_{-\infty}^{\infty} |u| f\left(u|x - \frac{L_n}{2}\right) du f_{u_x}\left(x - \frac{L_n}{2}\right) \right. \\ &\quad \left. + \int_{-\infty}^{\infty} |u| f\left(u|x + \frac{(m+1)L_n}{2m}\right) du f_{u_x}\left(x + \frac{(m+1)L_n}{2m}\right) \right\} \\ &= O(L_n^{-v-1}/\beta_n) = O(L_n^{-v}m^{-1}),\end{aligned}$$

where $f(\cdot|\cdot)$ is the conditional p.d.f. of u_i given u_{xi} . This completes the proof. ■

Lemma 5.2 Let $\Theta_{jn} = \frac{L_n}{n} \sum_{i=1}^n (X_i - x)^j K_h(X_i - x)$ for $j = 0, 1, 2$. Suppose A1-A6 hold. Then

- (i) $E\Theta_{0n} = 1 + o(1)$,
- (ii) $E\Theta_{1n} = o(h^2)$,
- (iii) $E\Theta_{2n} = h^2\mu_2(K) + o(h^2)$,
- (iv) $\text{Var}(\Theta_{jn}) = (h^{2j-1}L_n/n) \int z^{2j} K^2(z) dz + o(h^{2j-1}L_n/n)$ for $j = 0, 1, 2$.

Proof. Note that

$$\begin{aligned}E\Theta_{jn} &= \frac{L_n}{2m+1} \sum_{\alpha=-m}^m E\left[(\mu_\alpha + u_{xi} - x)^j K_h(\mu_\alpha + u_{xi} - x)\right] \\ &= \frac{L_n h^j}{2m+1} \sum_{\alpha=-m}^m \int z^j K(z) f_{u_x}(x - \mu_\alpha + hz) dz.\end{aligned}$$

If $j = 0$, we have

$$E\Theta_{0n} = \frac{L_n}{2m+1} \sum_{\alpha=-m}^m f_{u_x}(x - \mu_\alpha) + \frac{L_n h^2 \mu_2(K)}{2m+1} \sum_{\alpha=-m}^m f_{u_x}''(x - \mu_\alpha) + R_{0n},$$

where $R_{0n} = \frac{L_n}{2(2m+1)} h^2 \int z^2 K(z) \int_0^1 \sum_{a=-m}^m [f''_{u_x}(x - \mu_\alpha + whz) - f''_{u_x}(x - \mu_\alpha)] (1-w) dw dz$ is the remainder term. Noting

$$\begin{aligned} \frac{L_n}{2m+1} \sum_{\alpha=-m}^m f_{u_x}(x - \mu_\alpha) &\approx \int_{-L_n/2}^{L_n/2} f_{u_x}(x-p) dp \rightarrow \int_{-\infty}^{\infty} f_{u_x}(x-p) dp = 1, \\ \frac{L_n}{2m+1} \sum_{\alpha=-m}^m |f''_{u_x}(x - \mu_\alpha)| &\approx \int_{-L_n/2}^{L_n/2} |f''_{u_x}(x-p)| dp \rightarrow \int_{-\infty}^{\infty} |f''_{u_x}(x-p)| dp < \infty, \end{aligned}$$

it follows that $R_{0n} = o(h^2)$ by the dominated convergence theorem. Thus, $E\Theta_{0n} = 1 + o(1)$. Similarly, for $j = 1, 2$, we have

$$E\Theta_{1n} = \frac{L_n h^2 \mu_2(K)}{2m+1} \sum_{\alpha=-m}^m f'_{u_x}(x - \mu_\alpha) + o(h^2) = h^2 \mu_2(K) \int_{-\infty}^{\infty} f'_{u_x}(x-p) dp + o(h^2) = o(h^2),$$

since $\int_{-\infty}^{\infty} f'_{u_x}(x-p) dp = 0$, and

$$E\Theta_{2n} = \frac{L_n h^2 \mu_2(K)}{2m+1} \sum_{\alpha=-m}^m f_{u_x}(x - \mu_\alpha) + o(h^4) = h^2 \mu_2(K) + o(h^2).$$

Now, by Jensen's inequality, a change of variables, and since (u_i, u_{xi}) is *iid*, we have

$$\begin{aligned} \text{Var}(\Theta_{jn}) &= \frac{L_n^2}{M(2m+1)^2} \sum_{\alpha=-m}^m \text{Var}\left((\mu_\alpha + u_x - x)^j K_h(\mu_\alpha + u_x - x)\right) \\ &\leq \frac{L_n^2 h^{2j}}{M(2m+1)^2 h} \sum_{\alpha=-m}^m \int z^{2j} K^2(z) f_{u_x}(x - \mu_\alpha + hz) dz \\ &= \frac{L_n h^{2j-1} \int z^{2j} K^2(z) dz}{M(2m+1)} \frac{L_n}{2m+1} \sum_{\alpha=-m}^m f_{u_x}(x - \mu_\alpha) + O\left(\frac{L_n h^{2j+1}}{M(2m+1)}\right) \\ &= \frac{L_n h^{2j-1} \int z^{2j} K^2(z) dz}{n} + o\left(\frac{L_n h^{2j-1}}{n}\right). \end{aligned}$$

This completes the proof of the lemma. ■

To prove Theorem 2.1, consider the usual bias-variance decomposition of $\hat{g}(x) - g(x)$:

$$\hat{g}(x) - g(x) = \frac{\frac{L_n}{n} \sum_{i=1}^n \{g(X_i) - g(x)\} K_h(X_i - x)}{\frac{L_n}{n} \sum_{i=1}^n K_h(X_i - x)} + \frac{\frac{L_n}{n} \sum_{i=1}^n u_i K_h(X_i - x)}{\frac{L_n}{n} \sum_{i=1}^n K_h(X_i - x)}. \quad (5.2)$$

By Lemmas 5.2(i) and (iv), the Chebyshev inequality, and Assumption A6,

$$\Theta_{0n} = \frac{L_n}{n} \sum_{i=1}^n K_h(X_i - x) = 1 + O_P\left((nh/L_n)^{-1/2}\right) = 1 + o_P(1). \quad (5.3)$$

Now write

$$\begin{aligned}
& \frac{L_n}{n} \sum_{i=1}^n \{g(X_i) - g(x)\} K_h(X_i - x) \\
&= \frac{L_n g'(x)}{n} \sum_{i=1}^n (X_i - x) K_h(X_i - x) + \frac{L_n g''(x)}{2n} \sum_{i=1}^n (X_i - x)^2 K_h(X_i - x) + R_n(x) \\
&\equiv g'(x) \Theta_{1n} + \frac{g''(x)}{2} \Theta_{2n} + R_n(x),
\end{aligned}$$

where $R_n(x) = \frac{L_n}{2n} \sum_{i=1}^n [g''(X_i^*) - g''(x)] (X_i - x)^2 K_h(X_i - x)$, and X_i^* lies between X_i and x . Following the proof of Lemma 5.2, it is easy to show that $R_n(x) = o(h^2)$. Then by Lemma 5.2 (ii)-(iv), the Chebyshev inequality, and Assumption A6

$$\begin{aligned}
& \frac{L_n}{n} \sum_{i=1}^n \{g(X_i) - g(x)\} K_h(X_i - x) \\
&= h^2 \mu_2(K) \frac{g''(x)}{2} + o_P(h^2) + O_P\left(\sqrt{\frac{L_n h}{n}}\right) \\
&= h^2 \mu_2(K) \frac{g''(x)}{2} + o_P(h^2).
\end{aligned} \tag{5.4}$$

Let $\Theta_{3n} = \sqrt{L_n h / (2mM)} \sum_{i=1}^n u_i K_h(X_i - x)$. We show that $\Theta_{3n} - E\Theta_{3n} \rightarrow_d N(0, \sigma^2 \nu_2(K))$. Write $\Theta_{3n} = \sum_{i=1}^n Z_i$, where $Z_i = \sqrt{L_n h / (2mM)} u_i K_h(X_i - x)$. Recall $\beta_n \equiv 2m/L_n$. By a change of variables, the Fubini theorem, Lemma 5.1 and under A2 and A6

$$\begin{aligned}
E\Theta_{3n} &= \frac{\sqrt{Mh}}{\sqrt{2m/L_n}} \sum_{\alpha=-m}^m E[u_i K_h(\mu_\alpha + u_{xi} - x)] \\
&= \frac{\sqrt{Mh\beta_n}}{\beta_n} \sum_{\alpha=-m}^m \int \int u K(z) f(u, x - \mu_\alpha + hz) dz du \\
&= \sqrt{Mh\beta_n} \int K(z) \left\{ \int u \frac{1}{\beta_n} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha + hz) du \right\} dz \\
&= \sqrt{Mh\beta_n} \int K(z) \left\{ O(\beta_n^{-2} + L_n^{-\nu}) + \int u f_u(u) du \right\} dz \\
&= \sqrt{nh/L_n} O(L_n^2/m^2 + L_n^{-\nu}) = o(1),
\end{aligned} \tag{5.5}$$

where we use the fact that the result in Lemma 5.1 also holds uniformly in a small neighborhood of x . Similarly, by (5.5) and Assumptions A1 and A2,

$$\begin{aligned}
\text{Var}(\Theta_{3n}) &= \frac{h}{\beta_n} \sum_{\alpha=-m}^m E[u_i^2 K_h^2(\mu_\alpha + u_{xi} - x)] - o(1) \\
&= \frac{1}{\beta_n} \sum_{\alpha=-m}^m \int u^2 K(z)^2 f(u, x - \mu_\alpha + hz) dz du + o(1)
\end{aligned}$$

$$= \nu_2(K) \int u^2 \int f(u, x-p) dp du + o(1) = \nu_2(K) \sigma^2 + o(1).$$

To show the asymptotic normality of $\Theta_{3n} - E\Theta_{3n}$, by the above variance calculation and the independence of $\{u_i, u_{xi}\}$ across i , it suffices to check the Liapounov condition. Let $\bar{Z}_i = Z_i - E(Z_i)$. Then by the C_r and Jensen inequalities and Assumptions A2 and A4-A6,

$$\begin{aligned} \sum_{i=1}^n E |\bar{Z}_i|^{2+\delta} &\leq 2^{2+\delta} \left(\frac{L_n h}{2mM} \right)^{1+\delta/2} \sum_{i=1}^n E |u_i K_h(X_i - x)|^{2+\delta} \\ &= 2^{2+\delta} \left(\frac{2mMh}{L_n} \right)^{-\delta/2} \frac{L_n}{2m} \sum_{\alpha=-m}^m \int |u|^{2+\delta} K(z)^{2+\delta} f(u, x - \mu_\alpha + hz) dz du \\ &\approx \left(\frac{2mMh}{L_n} \right)^{-\delta/2} 2^{2+\delta} \int K(z)^{2+\delta} dz \int |u|^{2+\delta} \frac{L_n}{2m} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha) du \\ &= \left(\frac{nh}{L_n} \right)^{-\delta/2} 2^{2+\delta} \int K(z)^{2+\delta} dz \left\{ E |u|^{2+\delta} + o(1) \right\} \rightarrow 0. \end{aligned}$$

By the Liapounov CLT,

$$\Theta_{3n} - E(\Theta_{3n}) \rightarrow_d N(0, \sigma^2 \nu_2(K)). \quad (5.6)$$

Combining (5.2)-(5.6) and noting that $n/(2mM) \rightarrow 1$ as $n \rightarrow \infty$, we obtain (2.5) and the proof is complete.

5.2 Proof of Theorem 2.2

The proof is analogous to that of Theorem 2.1, so only the differences are sketched here. First, under A3(i), A7, and A8, the result in Lemma 5.1 changes to

$$\int u \left[L(2m)^{-1} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha) - f_u(u) \right] du = O(m^{-2}). \quad (5.7)$$

To see this, noticing that $f(u, x - \frac{L}{2}) = f(u, x + \frac{(m+1)L}{2m}) = 0$ since there is zero density outside the support by A7 and A8, we have

$$\begin{aligned} &\frac{L}{2m} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha) - f_u(u) \\ &= \frac{L}{2m} \sum_{\alpha=-m}^m f\left(u, x - \frac{\alpha L}{2m}\right) - \int_{\underline{u}}^{\bar{u}} f(u, u_x) du_x \\ &= \left\{ \frac{L}{2m} \sum_{\alpha=-m}^m \frac{1}{2} \left[f\left(u, x - \frac{\alpha L}{2m}\right) + f\left(u, x - \frac{(\alpha+1)L}{2m}\right) \right] - \int_{-L/2}^{L/2} f(u, x-p) dp \right\} \\ &\quad + \left\{ \int_{-L/2}^{L/2} f(u, x-p) dp - \int_{\underline{u}}^{\bar{u}} f(u, u_x) du_x \right\} \\ &\equiv I_{1m}(u) + I_{2m}(u). \end{aligned}$$

Since $x + \frac{L}{2} > \bar{u}$ and $x - \frac{L}{2} < \underline{u}$, we have $I_{2m}(u) = \int_{x-L/2}^{x+L/2} f(u, u_x) du_x - \int_{\underline{u}}^{\bar{u}} f(u, u_x) du_x = 0$ by A7 and A8. By (5.1) and Assumption A3(i)

$$\begin{aligned} (2m/L)^2 \int_{-\infty}^{\infty} u I_{1m}(u) du &= \int_{-\infty}^{\infty} \frac{u}{12} \frac{L}{2m} \sum_{\alpha=-m}^m f_2''(u, x - p_\alpha) du \\ &\rightarrow \frac{1}{12} \int_{-\infty}^{\infty} u \int_{-L/2}^{L/2} f_2''(u, x - p) dp du = \frac{1}{12} \int_{-\infty}^{\infty} \int_{\underline{u}}^{\bar{u}} u f_2''(u, p) dp du, \end{aligned}$$

where p_α lies between μ_α and $\mu_{\alpha+1}$. Thus (5.7) holds.

Next, one can show that the results in Lemmas 5.2(i) and (iii)-(iv) hold with L_n replaced by L . The result in Lemma 5.2(ii) now becomes

$$\begin{aligned} E\Theta_{1n} &= \frac{h^2 \mu_2(K)}{2m/L} \sum_{\alpha=-m}^m f'_{u_x}(x - \mu_\alpha) + o(h^2) = h^2 \mu_2(K) \int_{-L/2}^{L/2} f'_{u_x}(x - p) dp + o(h^2) \\ &= h^2 \mu_2(K) \int_{x-L/2}^{x+L/2} f'_{u_x}(p) dp + o(h^2) = h^2 \mu_2(K) [f_{u_x}(\bar{u}) - f_{u_x}(\underline{u})] + o(h^2), \end{aligned}$$

where the dominant term vanishes if and only if $f_{u_x}(\bar{u}) = f_{u_x}(\underline{u})$.

As to Θ_{3n} defined in the proof of Theorem 2.1, we have that for sufficiently large n

$$\begin{aligned} E\Theta_{3n} &= \frac{\sqrt{Mh}}{\sqrt{2m/L}} \sum_{\alpha=-m}^m E[u_i K_h(\mu_\alpha + u_{xi} - x)] \\ &= \frac{\sqrt{2Mmh/L}}{2m/L} \sum_{\alpha=-m}^m E[u_i K_h(\mu_\alpha + u_{xi} - x)] \\ &= \sqrt{2Mmh/L} \int K(z) \left\{ \int u \frac{1}{2m/L} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha + hz) du \right\} dz \\ &= \sqrt{2Mmh/L} \int K(z) \left\{ O(m^{-2}) + \int u f_u(u) du \right\} dz \\ &= \sqrt{nh} O(m^{-2}) = o(1), \end{aligned}$$

where we have used the fact if $x + \frac{L}{2} > \bar{u}$ and $x - \frac{L}{2} < \underline{u}$, then $x + hz + \frac{L}{2} \geq \bar{u}$ and $x + hz - \frac{L}{2} \leq \underline{u}$ for sufficiently large n and fixed z as $h \rightarrow 0$.

These results imply that the bias and variance calculations in the proof of Theorem 2.1 continue to hold with the change due to $E\Theta_{1n}$ and with L_n replaced by L everywhere. In addition, the Liapounov condition is also satisfied. This completes the proof of the theorem.

5.3 Proof of Theorem 3.2

Noting that $\sqrt{n}(\hat{\theta} - \theta) = (n^{-1}\mathbf{X}'\mathbf{X})^{-1} n^{-1/2}\mathbf{X}'\mathbf{u}$, we have $\sqrt{n}D_n(\hat{\theta} - \theta - (\mathbf{X}'\mathbf{X})^{-1} E(\mathbf{X}'\mathbf{u})) = (D_n^{-1}n^{-1}\mathbf{X}'\mathbf{X}D_n^{-1})^{-1} D_n^{-1}n^{-1/2}(\mathbf{X}'\mathbf{u} - E(\mathbf{X}'\mathbf{u}))$. We prove the theorem by showing that

$$\Gamma_n \equiv D_n^{-1}n^{-1}\mathbf{X}'\mathbf{X}D_n^{-1} \rightarrow_p \lim_{n \rightarrow \infty} \begin{pmatrix} 1 & L_n^{-1}\mu_x \\ L_n^{-1}\mu_x & \frac{1}{12} + \frac{E(u_{xi}^2)}{L_n^2} \end{pmatrix} \equiv \Gamma, \quad (5.8)$$

and

$$A_n \equiv D_n^{-1} n^{-1/2} (\mathbf{X}\mathbf{u} - E(\mathbf{X}\mathbf{u})) \rightarrow_d N(0, \Omega), \quad (5.9)$$

where

$$\Omega = \lim_{n \rightarrow \infty} \begin{pmatrix} \sigma^2 & \frac{E(u_{x1}u_1^2)}{L_n} \\ \frac{E(u_{x1}u_1^2)}{L_n} & \frac{\sigma^2}{12} + \frac{\text{Var}(u_{x1}u_1)}{L_n^2} \end{pmatrix}.$$

First, by the fact that $\bar{X} = \bar{u}_x \rightarrow_p \mu_x$, and (3.3), we have

$$\begin{aligned} L_n^{-2} n^{-1} \sum_{i=1}^n X_i^2 &= L_n^{-2} n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2 + L_n^{-2} \bar{X}^2 \\ &= \frac{1}{12} + \frac{\sigma_x^2}{L_n^2} + \frac{\mu_x^2}{L_n^2} + o_P(1) = \frac{1}{12} + \frac{E(u_{x1}^2)}{L_n^2} + o_P(1). \end{aligned}$$

Thus (5.8) follows. To show (5.9), by the Cramér-Wold device it suffices to show that for any $\omega = (\omega_1, \omega_2)'$ with $\|\omega\| = 1$, we have

$$\omega' A_n = n^{-1/2} \sum_{i=1}^n \{\omega_1 u_i + \omega_2 L_n^{-1} [X_i u_i - E(X_i u_i)]\} \rightarrow_d N(0, \omega' \Omega \omega). \quad (5.10)$$

By construction, $E(\omega' A_n) = 0$. We now calculate the asymptotic variance of $\omega' A_n$:

$$\begin{aligned} \text{Var}(\omega' A_n) &= n^{-1} \sum_{i=1}^n \text{Var}(\omega_1 u_i + \omega_2 L_n^{-1} [X_i u_i - E(X_i u_i)]) \\ &= \omega_1^2 \sigma^2 + 2\omega_1 \omega_2 L_n^{-1} n^{-1} \sum_{i=1}^n E(X_i u_i^2) + \omega_2^2 L_n^{-2} n^{-1} \sum_{i=1}^n \text{Var}(X_i u_i) \\ &\rightarrow \omega' \Omega \omega, \end{aligned}$$

because $L_n^{-1} n^{-1} \sum_{i=1}^n E(X_i u_i^2) = L_n^{-1} E(u_{xi} u_i^2)$, and

$$\begin{aligned} \frac{1}{n L_n^2} \sum_{i=1}^n \text{Var}(X_i u_i) &= \frac{1}{(2m+1) L_n^2} \sum_{\alpha=-m}^m \frac{1}{M} \sum_{i \in A_\alpha} \text{Var}\left(\left(\frac{L_n \alpha}{2m} + u_{xi}\right) u_i\right) \\ &= \frac{1}{(2m+1) L_n^2} \sum_{\alpha=-m}^m \frac{1}{M} \sum_{i \in A_\alpha} \left\{ \frac{L_n^2 \alpha^2}{4m^2} \sigma^2 + \text{Var}(u_{xi} u_i) + \frac{L_n \alpha}{m} E[u_{xi} u_i^2] \right\} \\ &= \frac{\sigma^2}{(2m+1) 2m^2} \frac{m(m+1)(2m+1)}{6} + \frac{\text{Var}(u_{xi} u_i)}{L_n^2} \\ &\rightarrow \frac{\sigma^2}{12} + \lim_{n \rightarrow \infty} \frac{\text{Var}(u_{xi} u_i)}{L_n^2}. \end{aligned}$$

Let $\xi_{in} = n^{-1/2} \{ \omega_1 u_i + \omega_2 L_n^{-1} [X_i u_i - E(X_i u_i)] \}$. By the C_r inequality,

$$\begin{aligned}
\sum_{i=1}^n E |\xi_{in}|^{2+\delta} &= \frac{1}{n^{1+\delta/2}} \sum_{i=1}^n E |\omega_1 u_i + \omega_2 L_n^{-1} [X_i u_i - E(X_i u_i)]|^{2+\delta} \\
&\leq \frac{2^{1+\delta}}{n^{1+\delta/2}} \sum_{i=1}^n \left\{ E |\omega_1 u_i|^{2+\delta} + |\omega_2 L_n^{-1}|^{2+\delta} E \|X_i u_i - E(X_i u_i)\|^{2+\delta} \right\} \\
&\leq \frac{2^{1+\delta} |\omega_1|^{2+\delta}}{n^{1+\delta/2}} \sum_{i=1}^n E |u_i|^{2+\delta} + |\omega_2 L_n^{-1}|^{2+\delta} \frac{2^{3+2\delta}}{n^{1+\delta/2}} \sum_{i=1}^n E \|X_i u_i\|^{2+\delta} \\
&= O(n^{-\delta/2}) = o(1).
\end{aligned}$$

Then (5.10) follows by the Liapounov CLT.

5.4 Proof of Theorem 3.3

For notational simplicity, let $K_{ix} = K_h(X_i - x)$. Recall $b_n = c_n b$, $a_n = c_n a$, and in the special case where $L_n = L$ is fixed, $c_n = 1$. We first prove some lemmas that are used in the proof of Theorem 3.3.

Lemma 5.3 *Let $\Phi_{jn} = \int_{a_n}^{b_n} x^j \hat{f}(x) dx$ for $j = 0, 1, 2$. Suppose the conditions in Theorem 3.3 hold. Then for $j = 0, 1, 2$,*

$$\begin{aligned}
(i) \quad E\Phi_{jn} &= \frac{b_n^{j+1} - a_n^{j+1}}{j+1} + O(h^2 c_n^{j+1}), \\
(ii) \quad \text{Var}(\Phi_{jn}) &= O\left(\frac{L_n c_n^{2j+1}}{n}\right), \\
(iii) \quad c_n^{-(j+1)} \Phi_{jn} &= \frac{b^{j+1} - a^{j+1}}{j+1} + o(1).
\end{aligned}$$

Proof. Recall $N = n/L_n = (2m+1)M/L_n$. By the Fubini theorem and Taylor expansion,

$$\begin{aligned}
E\Phi_{jn} &= \frac{L_n}{2m+1} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int x^j K(z) f_{u_x}(x - \mu_\alpha + hz) dz dx \\
&= \frac{L_n}{2m+1} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} x^j f_{u_x}(x - \mu_\alpha) dx \\
&\quad + \frac{h^2 \mu_2(K) L_n}{2m+1} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} x^j f_{u_x}''(x - \mu_\alpha) dx + R_{jn}.
\end{aligned} \tag{5.11}$$

where $R_{jn} = \frac{h^2 L_n}{2(2m+1)} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int \int x^j z^2 K(z) [f_{u_x}''(x - \mu_\alpha + whz) - f_{u_x}''(x - \mu_\alpha)] (1-w) dw dz dx$ is the remainder term. Noting

$$\begin{aligned}
\frac{L_n}{2m+1} \sum_{\alpha=-m}^m f_{u_x}(x - \mu_\alpha) &\approx \int_{L_n/2}^{L_n/2} f_{u_x}(x - p) dp \rightarrow \int_{-\infty}^{\infty} f_{u_x}(x - p) dp = 1, \\
\frac{L_n}{2m+1} \sum_{\alpha=-m}^m f_{u_x}''(x - \mu_\alpha) &\approx \int_{L_n/2}^{L_n/2} f_{u_x}''(x - p) dp \rightarrow \int_{-\infty}^{\infty} f_{u_x}''(x - p) dp < \infty,
\end{aligned}$$

it follows that the first term in (5.11) is approximately

$$\int_{a_n}^{b_n} x^j dx = \frac{b_n^{j+1} - a_n^{j+1}}{j+1}, \quad (5.12)$$

the second term in (5.11) is approximately

$$h^2 \mu_2(K) \int_{-\infty}^{\infty} f_{u_x}''(p) dp \int_{a_n}^{b_n} x^j dx = O(h^2 c_n^{j+1}),$$

and $R_{jn} = o(h^2 c_n^{j+1})$ by dominated convergence.

Now, by Jensen's inequality, the Fubini theorem, a change of variables, and since (u_i, u_{xi}) is *iid*, we have

$$\begin{aligned} & \text{Var}(\Phi_{jn}) \\ &= \frac{L_n^2}{(2m+1)^2 M^2} \sum_{i=1}^n \text{Var} \left(\int_{a_n}^{b_n} x^j K_h(X_i - x) dx \right) \\ &= \frac{L_n^2}{(2m+1)^2 M} \sum_{\alpha=-m}^m \text{Var} \left(\int_{a_n}^{b_n} x^j K_h(\mu_\alpha + u_{x1} - x) dx \right) \\ &\leq \frac{L_n^2}{(2m+1)^2 M} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int_{a_n}^{b_n} x^j \tilde{x}^j E[K_h(\mu_\alpha + u_{x1} - x) K_h(\mu_\alpha + u_{x1} - \tilde{x})] dx d\tilde{x} \\ &= \frac{L_n^2}{(2m+1)^2 M} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int_{a_n}^{b_n} x^j \tilde{x}^j \int K_h(\mu_\alpha + u_{x1} - x) K_h(\mu_\alpha + u_{x1} - \tilde{x}) f_{u_x}(u_{x1}) du_{x1} dx d\tilde{x} \\ &= \frac{L_n^2}{h(2m+1)^2 M} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int_{a_n}^{b_n} x^j \tilde{x}^j \int K(z) K\left(z + \frac{x - \tilde{x}}{h}\right) f_{u_x}(x - \mu_\alpha + hz) dz dx d\tilde{x} \\ &\approx \frac{L_n}{nh} \int_{a_n}^{b_n} \int_{a_n}^{b_n} x^j \tilde{x}^j \int K(z) K\left(z + \frac{x - \tilde{x}}{h}\right) \frac{L_n}{2m+1} \sum_{\alpha=-m}^m f_{u_x}(x - \mu_\alpha) dz dx d\tilde{x} \\ &= \frac{L_n}{nh} \int_{a_n}^{b_n} \int_{a_n}^{b_n} x^j \tilde{x}^j \int K(z) K\left(z + \frac{x - \tilde{x}}{h}\right) dz dx d\tilde{x} \{1 + o(1)\} \\ &= \frac{L_n}{n} \int_{a_n}^{b_n} \int_{(a_n-x)/h}^{(b_n-x)/h} x^j (x + hv)^j \int K(z) K(z - v) dz dv dx \{1 + o(1)\} \\ &= O\left(\frac{L_n c_n^{2j+1}}{n}\right). \end{aligned}$$

(iii) follows from (i) and (ii) and the Chebyshev inequality. ■

Lemma 5.4 *Let $\Xi_{jn} = \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n x^j (X_i - x) K_{ix} dx$ for $j = 0, 1$. Then $\Xi_{jn} = O_P(h^2 \sqrt{n/L_n} c_n^{j+1} + h c_n^{j+1/2})$.*

Proof. By the Fubini theorem and Taylor expansion,

$$\begin{aligned}
E(\Xi_{jn}) &= \sqrt{\frac{ML_n}{2m+1}} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} E[x^j (\mu_\alpha + u_{x1} - x) K_h(\mu_\alpha + u_{x1} - x)] dx \\
&= \sqrt{\frac{ML_n}{2m+1}} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int x^j z K(z) f_{u_x}(x - \mu_\alpha + hz) dz dx \\
&= h^2 \sqrt{\frac{n}{L_n}} \mu_2(K) \left\{ \int_{a_n}^{b_n} x^j \frac{L_n}{2m+1} \sum_{\alpha=-m}^m f''_{u_x}(x - \mu_\alpha) dx \{1 + o(1)\} \right\} \\
&= h^2 \sqrt{n/L_n} \mu_2(K) \int_{a_n}^{b_n} x^j dx \int f''_{u_x}(p) dp \{1 + o(1)\} = O\left(h^2 \sqrt{n/L_n} c_n^{j+1}\right).
\end{aligned}$$

Analogous to the proof of Lemma 5.3(ii), we can show that $\text{Var}(\Xi_{jn}) = O(h^2 c_n^{2j+1})$. The result follows from the Chebyshev inequality. ■

To prove Theorem 3.3, noticing that if $g(x) = \beta_0 + \beta_1 x$, then

$$\begin{aligned}
&\int_{a_n}^{b_n} \hat{g}(x) \hat{f}(x) dx \\
&= \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^n K_h(X_i - x) y_i dx \\
&= \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^n K_{ix} [\beta_0 + \beta_1 x + \beta_1 (X_i - x) + u_i] dx \\
&= \beta_0 \int_{a_n}^{b_n} \hat{f}(x) dx + \beta_1 \int_{a_n}^{b_n} x \hat{f}(x) dx + \beta_1 \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^n (X_i - x) K_{ix} dx + \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^n K_{ix} u_i dx,
\end{aligned}$$

and similarly $\int_{a_n}^{b_n} x \hat{g}(x) \hat{f}(x) dx = \beta_0 \int_{a_n}^{b_n} x \hat{f}(x) dx + \beta_1 \int_{a_n}^{b_n} x^2 \hat{f}(x) dx + \beta_1 \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^n x (X_i - x) K_{ix} dx + \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^n x K_{ix} u_i dx$. It follows that

$$\begin{aligned}
&\sqrt{N} C_n (\tilde{\theta} - \theta) \\
&= \beta_1 (C_n^{-1} Q_n C_n^{-1})^{-1} C_n^{-1} \begin{pmatrix} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n (X_i - x) K_{ix} dx \\ \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n x (X_i - x) K_{ix} dx \end{pmatrix} \\
&+ (C_n^{-1} Q_n C_n^{-1})^{-1} C_n^{-1} \begin{pmatrix} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n K_{ix} u_i dx \\ \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n x K_{ix} u_i dx \end{pmatrix}. \tag{5.13}
\end{aligned}$$

By Lemma 5.3,

$$C_n^{-1} Q_n C_n^{-1} \rightarrow_p \begin{pmatrix} b-a & \frac{b^2-a^2}{2} \\ \frac{b^2-a^2}{2} & \frac{b^3-a^3}{3} \end{pmatrix} = Q, \tag{5.14}$$

where $\det(Q) = (b-a)^4/12 > 0$ as $b \neq a$. This, together with Lemma 5.4, implies that the first term in (5.13) is $O_P(h^2 \sqrt{n/L_n} c_n^{1/2} + h) = o_P(1)$ by Assumption A6*. We are left to show that the second term in (5.13) is asymptotically $N(0, \sigma^2 Q^{-1})$.

Let $\omega = (\omega_1, \omega_2)'$ be such that $\|\omega\| = 1$, and define

$$\Theta_n = \omega' C_n^{-1} \left(\begin{array}{c} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n K_{ix} u_i dx \\ \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n x K_{ix} u_i dx \end{array} \right) = N^{-1/2} \sum_{i=1}^n \int_{a_n}^{b_n} \left(c_n^{-1/2} \omega_1 + c_n^{-3/2} \omega_2 x \right) K_{ix} u_i dx.$$

We complete the proof by showing that $E(\Theta_n) = o(1)$ and

$$\Theta_n - E(\Theta_n) \rightarrow_d N(0, \sigma^2 \omega' Q \omega). \quad (5.15)$$

By a change of variables, the Fubini theorem, Lemma 5.1, and Assumptions A2, A3(i) and A6*, we obtain

$$\begin{aligned} E\Theta_n &= \sqrt{\frac{ML_n}{2m+1}} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} E \left[\left(c_n^{-1/2} \omega_1 + c_n^{-3/2} \omega_2 x \right) K_h(\mu_\alpha + u_{x1} - x) u_1 \right] dx \\ &= \sqrt{\frac{ML_n}{2m+1}} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int \int \left(c_n^{-1/2} \omega_1 + c_n^{-3/2} \omega_2 x \right) u K(z) f(u, x - \mu_\alpha + hz) dz du dx \\ &= \sqrt{\frac{n}{L_n}} \int_{a_n}^{b_n} \left(c_n^{-1/2} \omega_1 + c_n^{-3/2} \omega_2 x \right) \int K(z) \left\{ \int u \frac{L_n}{2m+1} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha + hz) du \right\} dz dx \\ &= \sqrt{N} \int_{a_n}^{b_n} \left(c_n^{-1/2} \omega_1 + c_n^{-3/2} \omega_2 x \right) \int K(z) \left\{ O((m/L_n)^{-2} + h^2) + \int u f(u) du \right\} dz dx \\ &= \sqrt{N} c_n^{1/2} O((m/L_n)^{-2} + h^2) = o(1). \end{aligned}$$

Next, letting $w_n(x) = c_n^{-1/2} \omega_1 + c_n^{-3/2} \omega_2 x$, we have

$$\begin{aligned} &\text{Var}(\Theta_n) \\ &= N^{-1} \sum_{i=1}^n \text{Var} \left(u_i \int_{a_n}^{b_n} \left(c_n^{-1/2} \omega_1 + c_n^{-3/2} \omega_2 x \right) K_{ix} dx \right) \\ &= \frac{L_n}{2m+1} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int_{a_n}^{b_n} E[w_n(x) w_n(\tilde{x}) u_1^2 K_h(\mu_\alpha + u_{x1} - x) K_h(\mu_\alpha + u_{x1} - \tilde{x})] dx d\tilde{x} + o(1) \\ &= \frac{L_n}{2m+1} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int_{a_n}^{b_n} w_n(x) w_n(\tilde{x}) \int \int u^2 K_h(\mu_\alpha + u_x - x) K_h(\mu_\alpha + u_x - \tilde{x}) f(u, u_x) du_x du \\ &\quad dx d\tilde{x} + o(1) \\ &= \frac{L_n}{2m+1} \sum_{\alpha=-m}^m \int_{a_n}^{b_n} \int_{a_n}^{b_n} w_n(x) w_n(\tilde{x}) \int \int u^2 K(z) K\left(z + \frac{x - \tilde{x}}{h}\right) f(u, x - \mu_\alpha + hz) dz du dx d\tilde{x} \\ &\quad + o(1) \\ &= \int_{a_n}^{b_n} \int_{a_n}^{b_n} w_n(x) w_n(\tilde{x}) \int \int u^2 K(z) K\left(z + \frac{x - \tilde{x}}{h}\right) \frac{L_n}{2m+1} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha) dz du dx d\tilde{x} + o(1) \end{aligned}$$

$$\begin{aligned}
&= \sigma^2 \int_{a_n}^{b_n} \int_{a_n}^{b_n} w_n(x) w_n(\tilde{x}) \int K(z) K\left(z + \frac{x - \tilde{x}}{h}\right) dz dx d\tilde{x} + o(1) \\
&= \sigma^2 \int_{a_n}^{b_n} \int_{(a_n-x)/h}^{(b_n-x)/h} w_n(x) w_n(x+hv) \int K(z) K(z-v) dz dv dx + o(1) \\
&= \sigma^2 \int_{a_n}^{b_n} \left(c_n^{-1/2} \omega_1 + c_n^{-3/2} \omega_2 x\right)^2 dx + o(1) \rightarrow \sigma^2 \omega' Q \omega,
\end{aligned}$$

where we have used the fact that $\frac{L_n}{2m} \sum_{\alpha=-m}^m f(u, x - \mu_\alpha) \rightarrow \int_{-\infty}^{\infty} f(u, x - p) dp = f(u)$. To show the asymptotic normality of $\Theta_n - E(\Theta_n)$, by the above variance calculation and the independence of (u_i, u_{xi}) across i , it suffices to check the Liapounov condition. Let $\bar{\xi}_{ni} = \xi_{ni} - E(\xi_{ni})$, where $\xi_{ni} = N^{-1/2} \int_{a_n}^{b_n} w_n(x) K_{ix} u_i dx$. Then by the C_r inequality,

$$\begin{aligned}
\sum_{i=1}^n E |\bar{\xi}_{ni}|^{2+\delta} &\leq 2^{1+\delta} N^{-(1+\delta/2)} \sum_{i=1}^n E \left| u_i \int_{a_n}^{b_n} w_n(x) K_{ix} dx \right|^{2+\delta} \\
&\quad + 2^{1+\delta} N^{-(1+\delta/2)} \sum_{i=1}^n \left| E \left[u_i \int_{a_n}^{b_n} w_n(x) K_{ix} dx \right] \right|^{2+\delta} \\
&\equiv L_{n1} + L_{n2}, \text{ say.}
\end{aligned}$$

First,

$$\begin{aligned}
L_{n1} &= 2^{1+\delta} N^{-\delta/2} \frac{L_n}{2m+1} \sum_{\alpha=-m}^m \int \int \left| u \int_{a_n}^{b_n} w_n(x) K_h(\mu_\alpha + u_x - x) dx \right|^{2+\delta} f(u, u_x) du_x du \\
&\leq c_\delta N^{-\delta/2} \frac{L_n}{2m+1} \sum_{\alpha=-m}^m \int \int |u|^{2+\delta} \left| \int_{a_n}^{b_n} K_h(\mu_\alpha + u_x - x) dx \right|^{2+\delta} f(u, u_x) du_x du \\
&\leq c_\delta N^{-\delta/2} \frac{L_n}{2m+1} \sum_{\alpha=-m}^m \int \int |u|^{2+\delta} f(u, u_x) du_x du \\
&= c_\delta N^{-\delta/2} E |u|^{2+\delta} L_n = O(N^{-\delta/2} L_n) = o(1),
\end{aligned}$$

where $c_\delta = 2^{1+\delta} \sup_{a_n \leq |x| \leq b_n} |w_n(x)|^{2+\delta} < \infty$. By the Jensen inequality, $L_{n2} \leq L_{n1} = o(1)$. Then by the Liapounov CLT, (5.15) follows, and the proof is complete.

5.5 Proof of Theorem 3.4

Let $B_n = n^{-1} \mathbf{X}' \mathbf{X}$. Noting that

$$\frac{1}{n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2} \begin{pmatrix} \bar{X} \tilde{c} \\ -\tilde{c} \end{pmatrix} = B_n^{-1} n^{-1} \sum_{i=1}^n \begin{pmatrix} 0 \\ -X_i \tilde{u}_i \end{pmatrix},$$

we have

$$\begin{aligned}
\sqrt{N}C_n(\hat{\theta}_c - \theta) &= \sqrt{N}C_n(\hat{\theta} - \theta) + C_n B_n^{-1} n^{-1} \sum_{i=1}^n \begin{pmatrix} 0 \\ -X_i \tilde{u}_i \end{pmatrix} \\
&= \sqrt{N}C_n B_n^{-1} n^{-1} \sum_{i=1}^n \begin{pmatrix} u_i \\ X_i u_i - X_i \tilde{u}_i \end{pmatrix} \\
&= \sqrt{N}C_n B_n^{-1} n^{-1} \sum_{i=1}^n \begin{pmatrix} u_i \\ (X_i, X_i^2) (\tilde{\theta} - \theta) \end{pmatrix}.
\end{aligned}$$

Let $\omega = (\omega_1, \omega_2)'$ with $\|\omega\| = 1$. Define

$$T_n = \sqrt{N} \omega' C_n B_n^{-1} n^{-1} \sum_{i=1}^n \begin{pmatrix} u_i \\ (X_i, X_i^2) (\tilde{\theta} - \theta) \end{pmatrix}.$$

By the Cramér-Wold device, it suffices to show that

$$T_n \rightarrow_d N(0, \omega' \Psi \omega). \quad (5.16)$$

We show (5.16) by distinguishing whether L_n is allowed to approach ∞ as $n \rightarrow \infty$.

Case 1. $L_n \rightarrow \infty$ as $n \rightarrow \infty$. By the proof of Theorem 3.3,

$$\begin{aligned}
\sqrt{N} c_n^{3/2} (\tilde{\beta}_1 - \beta_1) &= (q^{21}, q^{22}) \begin{pmatrix} c_n^{-1/2} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n K_{ix} u_i dx \\ c_n^{-3/2} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n x K_{ix} u_i dx \end{pmatrix} + o_P(1) \\
&= q^{21} c_n^{-1/2} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n K_{ix} u_i dx + q^{22} c_n^{-3/2} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^n x K_{ix} u_i dx + o_P(1).
\end{aligned}$$

where for $i, j = 1, 2$, q^{ij} is the (i, j) element of Q^{-1} :

$$Q^{-1} = \frac{12}{(b-a)^4} \begin{pmatrix} \frac{b^3-a^3}{3} & -\frac{b^2-a^2}{2} \\ -\frac{b^2-a^2}{2} & b-a \end{pmatrix} \equiv \begin{pmatrix} q^{11} & q^{12} \\ q^{21} & q^{22} \end{pmatrix}.$$

Noting that $\bar{X} = n^{-1} \sum_{i=1}^n u_{xi} \rightarrow_p \mu_x$, $L_n^{-2} n^{-1} \sum_{i=1}^n X_i^2 = L_n^{-2} n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2 + L_n^{-2} \bar{X}^2 \rightarrow_p \frac{1}{12}$, we have $S_{xnc}^2 / S_x^2 \rightarrow_p 1$ and $\bar{X} / S_x^2 \rightarrow_p 0$, where $S_x^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ and $S_{xnc}^2 = n^{-1} \sum_{i=1}^n X_i^2$. Also note that

$$B_n^{-1} = \frac{1}{S_x^2} \begin{pmatrix} S_{xnc}^2 & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix}. \quad (5.17)$$

It follows that

$$\begin{aligned}
T_n &= \sqrt{N} n^{-1} \omega' C_n \sum_{i=1}^n \frac{1}{S_x^2} \begin{pmatrix} S_{xnc}^2 u_i - \bar{X} (X_i, X_i^2) (\tilde{\theta} - \theta) \\ -\bar{X} u_i + (X_i, X_i^2) (\tilde{\theta} - \theta) \end{pmatrix} \\
&= \sqrt{N} n^{-1} \sum_{i=1}^n \frac{1}{S_x^2} \left\{ \left(\omega_2 c_n^{3/2} - \omega_1 c_n^{1/2} \bar{X} \right) (X_i, X_i^2) (\tilde{\theta} - \theta) + \left(\omega_1 c_n^{1/2} S_{xnc}^2 - \omega_2 c_n^{3/2} \bar{X} \right) u_i \right\} \\
&= \frac{\omega_2 c_n^{3/2} - \omega_1 c_n^{1/2} \bar{X}}{S_x^2} \left(c_n^{-1/2} \bar{X}, c_n^{-3/2} S_{xnc}^2 \right) \sqrt{N} C_n (\tilde{\theta} - \theta) \\
&\quad + L_n^{-1/2} \frac{1}{S_x^2} \left(\omega_1 c_n^{1/2} S_{xnc}^2 - \omega_2 c_n^{3/2} \bar{X} \right) n^{-1/2} \sum_{i=1}^n u_i \\
&= (0, \omega_2 - c_n^{-1} \omega_1 \mu_x) \sqrt{N} C_n (\tilde{\theta} - \theta) + \omega_1 (c_n/L_n)^{1/2} n^{-1/2} \sum_{i=1}^n u_i + o_p(1) \\
&= (\omega_2 - c_n^{-1} \omega_1 \mu_x) \sqrt{N} c_n^{3/2} (\tilde{\beta}_1 - \beta_1) + \omega_1 (c_n/L_n)^{1/2} n^{-1/2} \sum_{i=1}^n u_i + o_p(1) \\
&= (\omega_2 - c_n^{-1} \omega_1 \mu_x) \left\{ q^{21} c_n^{-1/2} N^{-1/2} \sum_{i=1}^n \int_{a_n}^{b_n} K_{ix} u_i dx + q^{22} c_n^{-3/2} N^{-1/2} \sum_{i=1}^n \int_{a_n}^{b_n} x K_{ix} u_i dx \right\} \\
&\quad + \omega_1 (c_n/L_n)^{1/2} n^{-1/2} \sum_{i=1}^n u_i + o_p(1) \\
&\rightarrow_d N \left(0, \sigma^2 \lim_{n \rightarrow \infty} \{ (\omega_2 - c_n^{-1} \omega_1 \mu_x)^2 q^{22} + \omega_1^2 (c_n/L_n) + 2 (\omega_2 - c_n^{-1} \omega_1 \mu_x) \omega_1 (c_n/L_n) \gamma \} \right) \\
&= N \left(0, \sigma^2 \lim_{n \rightarrow \infty} \omega' \begin{pmatrix} c_n^{-2} \mu_x^2 q^{22} + \frac{c_n}{L_n} & -c_n^{-1} \mu_x q^{22} + \frac{c_n}{L_n} \gamma \\ -c_n^{-1} \mu_x q^{22} + \frac{c_n}{L_n} \gamma & q^{22} \end{pmatrix} \omega \right),
\end{aligned}$$

where $\gamma = q^{21} (b - a) + \frac{1}{2} q^{22} (b^2 - a^2)$.

Case 2. $L_n = L$ is fixed as $n \rightarrow \infty$. In this case, $c_n = 1$ and we can write a_n and b_n as a and b throughout. By the proof of Theorem 3.3 (see (5.13) and arguments thereafter),

$$\begin{aligned}
\sqrt{N} (\tilde{\theta} - \theta) &= Q^{-1} N^{-1/2} \sum_{i=1}^n \begin{pmatrix} \int_a^b [K_{ix} u_i - E(K_{ix} u_i)] dx \\ \int_a^b x [K_{ix} u_i - E(K_{ix} u_i)] dx \end{pmatrix} + o_P(1) \\
&\equiv \sum_{i=1}^n \begin{pmatrix} \xi_{i1} \\ \xi_{i2} \end{pmatrix} + o_P(1),
\end{aligned}$$

where

$$\xi_{i1} = N^{-1/2} \left\{ q^{11} \int_a^b [K_{ix} u_i - E(K_{ix} u_i)] dx + q^{12} \int_a^b x [K_{ix} u_i - E(K_{ix} u_i)] dx \right\}, \quad (5.18)$$

and

$$\xi_{i2} = N^{-1/2} \left\{ q^{21} \int_a^b [K_{ix} u_i - E(K_{ix} u_i)] dx + q^{22} \int_a^b x [K_{ix} u_i - E(K_{ix} u_i)] dx \right\}. \quad (5.19)$$

Let

$$R_n = \sqrt{N}n^{-1} \sum_{i=1}^n \begin{pmatrix} u_i \\ (X_i, X_i^2) (\tilde{\theta} - \theta) \end{pmatrix}.$$

Then

$$\begin{aligned} \omega' R_n &= \omega_2 \sqrt{N} (\tilde{\beta}_1 - \beta_1) \left\{ n^{-1} \sum_{i=1}^n X_i^2 \right\} + \omega_2 \sqrt{N} (\tilde{\beta}_0 - \beta_0) \left\{ n^{-1} \sum_{i=1}^n X_i \right\} + \omega_1 \sqrt{N} n^{-1} \sum_{i=1}^n u_i \\ &= \omega_2 c_x \sqrt{N} (\tilde{\beta}_1 - \beta_1) + \omega_2 \mu_x \sqrt{N} (\tilde{\beta}_0 - \beta_0) + \omega_1 \sqrt{N} n^{-1} \sum_{i=1}^n u_i + o_P(1) \\ &= \sum_{i=1}^n \left(\omega_2 c_x \xi_{i2} + \omega_2 \mu_x \xi_{i1} + \omega_1 \sqrt{N} n^{-1} u_i \right) + o_P(1) \\ &\equiv \bar{R}_n + o_P(1), \end{aligned}$$

where recall $c_x = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n E(X_i^2) = \frac{L^2}{12} + E(u_{xi}^2)$. Let $c = (\mu_x, c_x)'$. Note that $E(\bar{R}_n) = 0$, and

$$\begin{aligned} &\text{Var}(\bar{R}_n) \\ &= \sum_{i=1}^n \text{Var} \left(\omega_2 c_x \xi_{i2} + \omega_2 \mu_x \xi_{i1} + \omega_1 \sqrt{N} n^{-1} u_i \right) \\ &= \omega_2^2 \sum_{i=1}^n \text{Var} (c_x \xi_{i2} + \mu_x \xi_{i1}) + 2\omega_2 \omega_1 \sum_{i=1}^n \text{Cov} (c_x \xi_{i2} + \mu_x \xi_{i1}, \sqrt{N} n^{-1} u_i) \\ &\quad + \omega_1^2 \sum_{i=1}^n \text{Var} (\sqrt{N} n^{-1} u_i) \\ &= \sigma^2 \omega_2^2 c' Q^{-1} c + 2\omega_2 \omega_1 \sigma^2 L^{-1} \left\{ q^{11} \int_a^b dx + q^{12} \int_a^b x dx + q^{21} \int_a^b dx + q^{22} \int_a^b x dx \right\} \\ &\quad + \omega_1^2 \sigma^2 L^{-1} + o(1) \\ &\rightarrow \sigma^2 \omega_2^2 c' Q^{-1} c + 2\omega_2 \omega_1 \sigma^2 L^{-1} + \omega_1^2 \sigma^2 L^{-1} = \sigma^2 \omega' \begin{pmatrix} L^{-1} & L^{-1} \\ L^{-1} & c' Q^{-1} c \end{pmatrix} \omega, \end{aligned}$$

where the third line follows from the definitions of ξ_{i1} and ξ_{i2} , Fubini, a change of variables,

and the following two limits:

$$\begin{aligned}
& \sum_{i=1}^n E \left(\xi_{i1} \sqrt{N} n^{-1} u_i \right) \\
&= n^{-1} \sum_{i=1}^n E \left\{ \left[q^{11} \int_a^b [K_{ix} u_i - E(K_{ix} u_i)] dx + q^{12} \int_a^b x [K_{ix} u_i - E(K_{ix} u_i)] dx \right] u_i \right\} \\
&= q^{11} n^{-1} \sum_{i=1}^n E \left\{ u_i^2 \int_a^b K_{ix} dx \right\} + q^{12} n^{-1} \sum_{i=1}^n E \left\{ u_i^2 \int_a^b x K_{ix} dx \right\} \\
&= q^{11} (2m+1)^{-1} \sum_{\alpha=-m}^m \int_a^b \int \int u^2 K(z) f(u, x - u_\alpha + hz) dz du dx \\
&\quad + q^{12} (2m+1)^{-1} \sum_{\alpha=-m}^m \int_a^b x \int \int u^2 K(z) f(u, x - u_\alpha + hz) dz du dx \\
&= q^{11} L^{-1} \int_a^b \int u^2 \left\{ \frac{L}{2m} \sum_{\alpha=-m}^m f(u, x - u_\alpha) \right\} du dx \{1 + o(1)\} \\
&\quad + q^{12} L^{-1} \int_a^b x \int u^2 \left\{ \frac{L}{2m} \sum_{\alpha=-m}^m f(u, x - u_\alpha) \right\} dx \{1 + o(1)\} \\
&= \sigma^2 q^{11} L^{-1} \int_a^b dx \{1 + o(1)\} + \sigma^2 q^{12} L^{-1} \int_a^b x dx \{1 + o(1)\} \\
&\rightarrow \sigma^2 L^{-1} \left\{ q^{11} \int_a^b dx + q^{12} \int_a^b x dx \right\};
\end{aligned}$$

and, in a similar way,

$$\begin{aligned}
\sum_{i=1}^n E \left(\xi_{i2} \sqrt{N} n^{-1} u_i \right) &= n^{-1} \sum_{i=1}^n E \left[q^{21} u_i^2 \int_a^b K_{ix} u_i dx + q^{22} u_i^2 \int_a^b x K_{ix} dx \right] \\
&\rightarrow \sigma^2 L^{-1} \left\{ q^{21} \int_a^b dx + q^{22} \int_a^b x dx \right\}.
\end{aligned}$$

The Liapounov condition follows from the verification in the proof of Theorem 3.3, the fact that $\sqrt{N} n^{-1} \sum_{i=1}^n u_i$ also satisfies the Liapounov condition, and the C_r inequality. It follows that

$$R_n \rightarrow_d N \left(0, \sigma^2 \begin{pmatrix} L^{-1} & L^{-1} \\ L^{-1} & c' Q^{-1} c \end{pmatrix} \right)$$

and

$$T_n = \omega' B_n^{-1} R_n \rightarrow_d N \left(0, \sigma^2 \omega' B^{-1} \begin{pmatrix} L^{-1} & L^{-1} \\ L^{-1} & c' Q^{-1} c \end{pmatrix} B^{-1} \omega \right).$$

References

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics* 2, 123-147.

- Carrasco, M., J-P. Florens, and E. Renault (2007). Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. *Handbook of Econometrics*, Vol. 6, Part 2, 5633 - 5751.
- Chen, L-A., A. H. Welsh, and W. Chan (2001). Estimators for the linear regression model based on winsorized observations. *Statistica Sinica* 11, 147-172.
- Cristóbal, J. A. C., P. F. Roca and W. G. Manteiga (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Annals of Statistics* 15, 603-609.
- Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*, vol. 66 of Monographs on Statistics and Applied Probability. Chapman and Hall.
- Florens, J-P. (2003). Inverse problems and structural econometrics: The example of instrumental variables. In *Advances in Economics and Econometrics: Theory and Applications*. (M. Dewatripont, L. P. Hansen and S. J. Turnovsky, eds.), 2, 284-311. Cambridge University Press.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Phillips, P. C. B. and L. Su (2009). Nonparametric structural estimation via continuous location shifts in an endogenous regressor. *Mimeo*, Dept. of Economics, Yale University.
- Simoni, A. (2009). Bayesian nonparametric estimation of asset price functionals. *Working paper*, Toulouse School of Economics.
- Stoer, J. and R. Bulirsch (1993). *Introduction to Numerical Analysis*. Springer-Verlag, New York.
- Wang, Q. and P. C. B. Phillips (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710-738.
- Wang, Q. and P. C. B. Phillips (2009b). Structural nonparametric cointegrating regression. *Econometrica* 77, 1901-1948.
- Wang, Q. and P. C. B. Phillips (2010). Asymptotic theory for zero energy density estimation with nonparametric regression applications. *Econometric Theory* (forthcoming).
- Welsh, A. H. (1987). The trimmed mean in the linear model. *Annals of Statistics* 15, 20-36.